Double toil and trouble: grade retention and academic performance in Spain

Álvaro Choi[a], María Gil[b], Mauro Mediavilla[c] & Javier Valbuena[d]

[a] *Universitat de Barcelona & IEB*

[b] *Universidad Autónoma de Madrid*

[c] *Universitat de València & IEB*

[d] *Universitat de Girona*

Correspondence author: María Gil Izquierdo. Department of Applied Economics. Faculty of Economics and Business. Universidad Autónoma de Madrid. Cantoblanco.28049, Madrid. Email: maria.gil@uam.es

Abstract

Grade retention is widely applied in the Spanish educational system, where almost 1 out of every 3 students have repeated at least one course by age 16. In that sense, Spain can be classified among those countries with a comprehensive educational system, where grade retention is applied as the main policy for levelling the performance of students. However, part of the variation in the rates of grade retention at an international level may also be explained by social believes and cultural factors. Several arguments are used for defending the use of grade retention. First, it provides students some extra time for maturing; second, it is a policy which is applied equally to all students; third, it may be a deterrent to low performance; finally, it may enhance overall performance as it transmits students a culture of effort. Those who stand against grade retention emphasize its inefficacy, its high cost and its negative impact on the student's level of motivation and may also generate discipline issues in schools. However, while most of the empirical evidence available for other countries shows the scarce effectiveness of grade retention for enhancing academic performance, this exercise is still to be performed for Spain.

Most studies analyzing the Spanish case cannot estimate precisely the effect of grade retention on academic achievement due to the lack of data rich enough for tackling methodological issues such as reverse causation. In this paper we attempt to overcome this limitation creating a pseudo-panel which combines microdata from PIRLS-2006 and PISA-2012. Our study provides evidence about the negative impact of grade retention and the importance of previous achievement. Therefore, this study underlines the importance of early intervention at primary school, or even before, in order to identify students at risk of grade retention.

Keywords: grade retention; academic achievement; PISA; PIRLS.

Acknowledgements: This research has received financial support from the Areces Foundation.

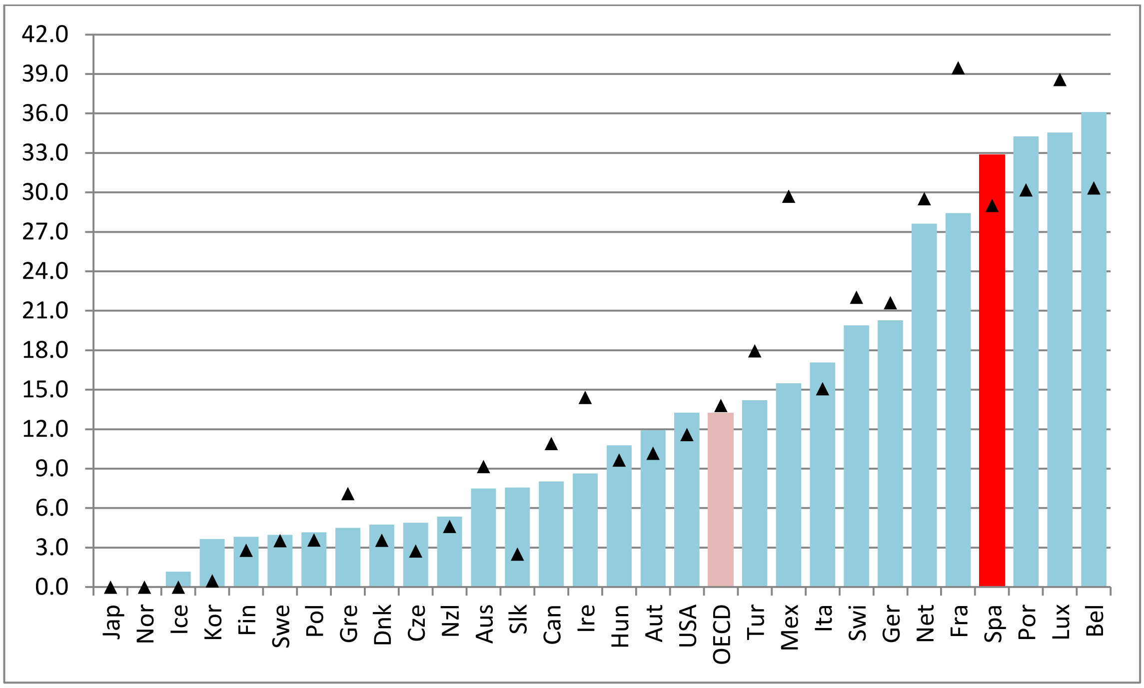
1 Grade retention as a universal solution

Grade retention is the educational practice that consists on making a student repeat an academic course. The main reason for applying this policy is to *punish* the students' poor performance for not being able to pass a certain number of subjects established in the syllabus. Several arguments have been used for defending its use. First, it provides students some extra time for maturing; second, it is a policy that is applied equally to all students; finally, it may enhance overall performance as it transmits students a culture of effort. In that sense, it may act as a deterrent to low performance (Manacorda, 2012). Nevertheless, grade retention is a controversial measure.

Those who stand against grade retention emphasize its inefficacy (Jimerson et al., 2002), its high cost (OECD, 2011) and its negative impact on the student's level of motivation, as repeaters are separated from their friends, have to repeat both the subjects they failed and those they passed, and may even be stigmatized as "slow" students (Martin, 2011). Besides, this may also generate discipline issues in schools (Crothers et al., 2010).

The application of grade retention is heterogeneous among educational systems (Figure 1): while countries such as Japan or Norway prefer social promotion, other countries such as Belgium or Portugal use grade retention intensively. Tradition, social believes and cultural factors seem to play an important role in explaining those differences (Goos et al., 2013).

Figure 1. Percentage of students who have repeated at least one course at age 15. OECD countries, PISA 2012.

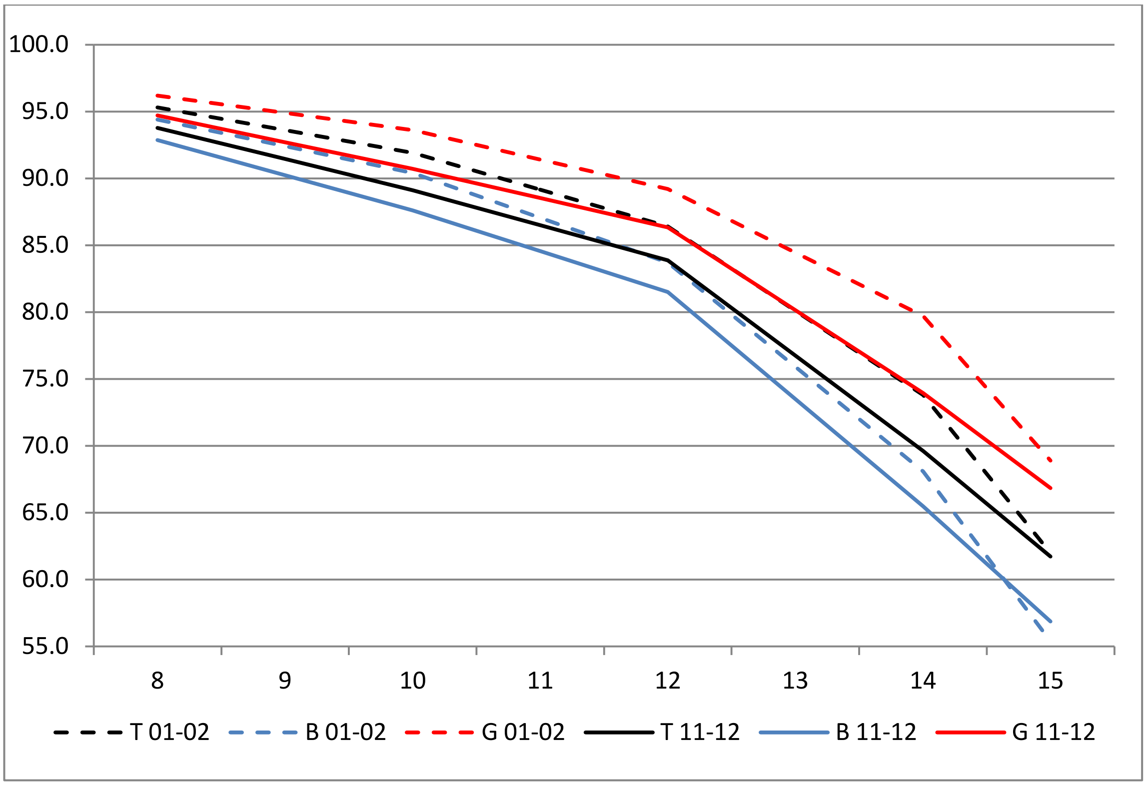


Source: self-elaboration from OECD data.

Note: ▲, repetition rate in year 2003.

Grade retention is widely applied in the Spanish educational system, where 1 out of every 3 students have repeated at least one course by age 16. Spain can be classified, according to Dupriez et al. (2008), among those countries with a comprehensive educational system, where grade retention is applied as the main policy for levelling the performance of students. Figure 2 provides specific information for the Spanish case.

Figure 2. Percentage of students in the theoretical age for each course. Academic year: 2011-12.



Source: self-elaboration from MECD (2014 –table C2). T: total; B: boys; G: girls.

Figure 2 shows repetition is mainly applied during compulsory secondary education[1] (ages 12 to 16) and is more common among Spanish boys. This seems to be consistent with their higher early school dropout rates (26% boys vs. 18.5% girls), as studies such as Jacob and Lefgren (2009) and Ou and Reynolds (2010) describe a positive relationship between grade

[1] Education in Spain is compulsory from ages 6 to 16. This comprises six years of primary school and four years of lower secondary education. Although not being compulsory, education from ages 3 to 5 is free -in public and private publicly-funded schools- and nearly universal. According to the Spanish Education Act (Ley Orgánica de Educación –LOE-, art. 20.4), students may only repeat once during primary school, and twice during the lower secondary education (ESO). Teachers decide which students should repeat, the main criteria established in the law being not passing three or more subjects. However, teachers are allowed to promote students who have not passed three subjects. Although the LOE is a national law, there are important differences in repetition rates among and within the Spanish regions (Comunidades Autónomas).

repetition and early school dropout. Anyway, although early school dropout figures have fallen since 2008 (mainly due to the economic crisis and the high youth unemployment rates) both Spanish girls and boys were still, in 2014, well above the EU 2020 Strategy benchmark (10% -15%, for the Spanish case-). Additionally, Figure 1 shows the number of repeaters in Spain increased between 2003 and 2012.

Grade retention policies come with substantial consequences for both students and schools. On the one hand, it is assumed that repetition can improve academic performance of low achievers by exposing them to an additional year of teaching so they catch up on the curriculum requirements. On the other hand, opponents to grade retention argue that it negatively impacts the student by reducing their level of motivation and stigmatizing them imposing new peer relationships. Therefore, it might negatively affect academic performance and increase the probability of dropping out of school (Holmes, 1989). Besides, holding back students in the same grade is a costly policy.

At an international level, there exists a vast literature within the economics of education that is concerned with grade retention policies. Still, the results are inconclusive ranging from substantial negative effects on academic achievement, to null grade retention effects or even positive (Jacob and Lefgren 2004). In a recent meta-analysis, Allen et. al (2009) highlight the crucial role of selection bias in determining the results of short-term grade retention effects. In the absence of pre-retention measures of academic ability results show a much larger negative association between repetition and academic performance. Besides, there is a growing literature on estimating the causal effect of grade retention policies on academic achievement (Dong, 2009; Eide and Showalter, 2001; Glick and Sahn, 2010; Gomes-Neto and Hanushek, 1994; Jacob and Lefgren, 2004, 2009 and Manacorda, 2012). Empirical findings are also mixed showing both negative and positive effects of repetition on academic performance and school dropout.

Compared to the large body of research at an international level, empirical research into the effect of grade repetition is fairly sparse in Spain. Indeed, there have been only a few recent empirical studies that look at the effect of grade retention. A common pattern in the results of the Spanish literature presented is as follows: grade retention shows a negative association with academic performance and increases the probability of school failure -dropping out of school before completing compulsory education (Calero et al., 2010; Calero and Escardíbul, 2007; Choi and Calero, 2013; Cordero et al., 2010; Guío and Choi, 2014; Mancebón et al, 2012; Salinas and Santín, 2012). The main limitation present in the results of these studies is that they cannot distinguish whether retention is the direct cause of poor academic achievement/school failure, or the result of previous characteristics associated to the students that increase their probability of failure. In the latter case, repetition would just be signalling those students at higher risk of academic failure (i.e. reverse causation). To the best of our knowledge, only the study by García-Pérez et al. (2014) do control for endogeneity using a Switching Regression Model, but their results crucially depend on endogenous selection.

Most studies analysing the Spanish case cannot estimate precisely the effect of grade retention on academic achievement. This is mainly due to the lack of data rich enough (i.e. longitudinal data) for tackling methodological issues such as reverse causation and endogenous selection. Therefore, there is a lack of robust empirical evidence for Spain. The aim of our study is twofold: to estimate robust results for the Spanish case and to propose a novel empirical approach of interest to comparative education researchers.

In this paper we attempt to overcome the limitations commented above by creating a pseudo-panel that combines micro-data from PIRLS-2006 and PISA-2012. Then, we study the role of previous achievement on grade retention -i.e. grade retention is usually a consequence of a cumulative process. Our analysis underlines the importance of early intervention at primary school, or even before, in order to identify students at risk of grade retention.

The remainder of the paper proceeds as follows. Section 2 discusses the data, methodology and their relative merits. Section 3 outlines relevant features of the results. Section 4 concludes.

2 Data and methodology

2.1 Introduction

It is common for researchers in social sciences to face some information restrictions when dealing with some specific issues, although databases are becoming more and more complete and thorough. This situation can be more severe in countries where data are not exhaustive enough to allow certain empirical analyses that are of great interest for the scientific community and policymakers alike. The analysis of grade retention in Spain is a good example of this situation. On the one hand, grade retention is widely applied in the Spanish educational system. On the other hand, there are no databases offering enough information to assess this issue.

As noted above, the lack of longitudinal data for Spain does not allow addressing some methodological considerations, such as reverse causation (i.e. do students have a low performance and, subsequently, have to repeat a year or do repeaters have a lower performance *because* of the negative impact of grade retention?). Therefore, as it has been seen in the previous chapter, most studies analysing the Spanish case cannot estimate precisely the effect of grade retention on academic achievement. In this paper, we attempt to overcome this limitation creating a pseudo-panel by merging microdata from two cross section international databases. In this chapter, first, we explain the two databases (2.1) and second, the methodology applied (2.2).

2.2 Data

There are several advantages of longitudinal data compared with either purely cross-sectional or purely time-series data. The two more important advantages are the ability to study dynamic relationships and to model the differences or heterogeneity, among subjects. The lack of this type of database for Spain forces us to surpass this limitation creating a pseudo-panel that combines microdata from two international cross section databases, PIRLS-2006 and PISA-2012.

The PISA 2012 (Programme for International Student Assessment) assesses on a triennial basis the extent to which 15-year-old students have acquired key competencies and skills. The assessment, which focuses on reading, mathematics, science and problem-solving, does not just assess whether students can reproduce what they have learned; it also examines how well they can extrapolate what they have learned and apply that knowledge in unfamiliar settings. This approach reflects the fact that modern societies reward individuals not for what they know, but for what they can do with what they know. 34 OECD member countries and 31 partner countries participated in the PISA 2012 assessment (OECD, 2014). PISA does not follow, however, the evolution of students along time and does not provide any information on their previous achievement. Given the existence of reverse causality issues, the direct estimation of an educational production function attempting to measure the impact of grade retention would therefore be biased. Still, this bias could be reduced controlling for prior academic performance.

We therefore turn to PIRLS-2006 data, our auxiliary sample. The PIRLS 2006 (Progress in International Reading Literacy Study) database was the second in the PIRLS series of studies carried out by the IEA (International Association for the Evaluation of Educational Achievement). Inaugurated in 2001 and conducted every 5 years, PIRLS is IEA’s assessment of students’ reading achievement at fourth grade. PIRLS 2006, implemented in 40 countries, assessed a range of reading comprehension processes within two major reading purposes—literary and informational of students aged 9/10. Moreover, this database provides information on their individual, household and school characteristics (Mullis et al., 2007). Interestingly, most students participating in PIRLS-2006 were born during 1996 and therefore belong to the same cohort as PISA-2012 students.

2.3 Methodology

In this paper we will focus on the reading skills of students aged 15/16 (PISA-2012) taking into account their predicted score at age 9/10. This last information will be extracted from the PIRLS-2006 database. This original approach is possible because both international assessments provide relevant information for students from the same cohort. Moreover, both assessments are regarded to be representative at the national level and share similar sampling designs and response rates. Then, we are able to identify a set of individual and household level variables present in both databases that are relevant for estimating academic performance. This allows us to apply a parametric approach for merging PIRLS-2006 results to the PISA-2012 database.

Similar approaches have been used in other fields of study. For instance, the use of the two samples two stage least squares (TSTSLS) methodology has been widely used in the intergenerational mobility literature[2]. In our study, we draw on the PISA-2012 data as the final objective is the identification of the impact of grade retention during lower secondary school on the academic performance of 15/16 years old students.

2.3.1 Imputing PIRLS 2006

The existence of a non-despicable amount of missing values becomes an issue when dealing with PIRLS. Consequently, before using this source of information as a donor for the main database, we need to *clean* the variables we intend to use for the analysis. In order to do so, we will use the technique of multiple imputation.

Although this methodology is known since the 1970s (Rubin, 1976), its development and implementation has been extended in recent years (Acock, 2005; Royston, 2005; Rubin, 1996; Sterne et al, 2009; van Buuren et al., 1999). This technique, with stochastic characteristics, allows to make full use of the data, obtain unbiased estimators and reflect the uncertainty that the non-response introduced in the estimation of parameters and, finally, preserve the dispersion of the distribution of the imputed variable (Rubin, 1996). Its implementation is based on replacing the unobserved data for m>1 possible simulated values from the maximum incorporation of predictive variables with missing values (Schafer, 1999; White et al., 2011).

The applicability of this method has been enhanced, in its general outline, with Monte Carlo methods based on Markov chains, known as MICE (Multiple imputation by chained equations) algorithms. Besides, multiple imputation is considered a flexible methodology to work with multivariate data and monotonous or arbitrary distributions of missing values.

[2] Jerrim et al. (2014) provide a recent review.

Finally, its suitability requires that the distribution pattern of missing values is random (MCAR – Missing Completely at Random- or MAR – Missing at Random -).

The PIRLS 2006 database for Spain consists of 4,094 observations. For the purpose of the analysis, we consider only students born in 1996. This results in a final sample of 3,771 observations. Missing values issues were also detected. For some observations there is no information on child schooling and education/occupation of the parents. Since this information is relevant to our investigation, they were eliminated from the sample (listwise deletion), checking afterwards that the resultant database kept similar average values for all the variables. The descriptive analysis of the resulting database (2,391 observations) is shown in Table 1.

Table 1. Missing Values Analysis

	N	*Missing values*	% *Missing*
Gender	2,389		0.08
Born in country	2,353	38	1.59
Father born in country	2,295	96	4.02
Mother born in country	2,308	83	3.47
ISCED O (in years)	2,338	44	1.84
Language of test at home	2,258	133	5.56
PC at home	2,348	43	1.80
Own room at home	2,341	50	2.09
Books at home	2,391	0	0.00
Parents highest education	2,287	104	4.35
Employment situation father	2,188	203	8.49
Employment situation mother	2,040	351	14.68
Parents highest occupation	2,228	163	6.82
Date of student birth (months)	2,391	0	0.00

Source: Own elaboration from PIRLS, 2006

As a preliminary step for imputation, we check the randomness of missing values using the dichotomized test correlations and checking the absence of outliers from box plots and histograms. We apply the multiple imputation technique using the MICE algorithm, systematized for the STATA software (version 13) using the command *my impute chained* (Royston and White, 2011; StataCorp, 2013). This estimation fills in missing values in multiple variables iteratively by using chained equations, a sequence of univariate imputation methods with fully conditional specification of prediction equations. By default, this command checks whether imputation variables have a monotone missing-data pattern and, if they do, imputes them using the monotone method. The imputation method applies possible values generated from a series of univariate models in which one variable is imputed based on a set of variables.

The MICE method is implemented in the chained method and uses a Gibbs-like algorithm to impute multiple variables sequentially using univariate fully conditional specifications. Finally, the estimation imputes the variables in a specific order -from the most observed to the least observed-. In this case, following the recommendation of Rubin (1996) and Acock (2005), we use all the available variables in the model to estimate unobserved data from three different empirical approaches (logit, ordered logit and multinomial logit) according to the particular characteristics of each variable. After the imputation process, we have a final database with 40 complete simulated databases.

2.3.2 Merging PIRLS and PISA data: creation of a pseudo-panel

Several methods were considered for linking PIRLS 2006 scores to PISA 2012. Semi-parametric matching techniques, using propensity score matching (PSM) and closest neighbour criteria were explored in first place. However, results obtained using PSM were unsatisfactory due to the different distributions in the characteristics of the individuals in both samples. We therefore moved to a parametric approach and proceeded as follows: we first estimated an educational production function using the PIRLS -auxiliary- database, the independent variables being those individual and household level variables also available at the PISA -main- sample[3]. We then apply the parameters obtained in the former regression to the PISA sample and obtain the predicted value that a student in the PISA database would have obtained in PIRLS. We therefore add an additional column to the PISA 2012 database: the predicted score of the student in PIRLS 2006. This procedure is repeated five times for each plausible value in PIRLS. As it may be seen, this linking technique follows a similar philosophy to the two-sample two-stage least squares (TSTSLS) technique.

2.3.3 Hierarchical linear model

PISA designs its sample using a two-stage method. In the first stage, a sample of schools is randomly selected from the whole list of centres that provide schooling for 15 year old students. In the second stage, a random sample of 35 students is chosen within each of the schools selected in the first stage. The probability of a school of being selected in PISA is proportional to its size. As a consequence, larger centres have a higher probability of being chosen; nevertheless, students in larger schools have smaller probabilities of being selected in comparison with students enrolled in smaller schools that have been chosen in the PISA sample. Therefore, the principle of independence of variables among the students of each centre does

[3] PSM results and results from this auxiliary regression are available upon request.

not hold, as students enrolled in the same school usually share socio-economic circumstances which make the average correlation among the variables of the students within the centre to be higher than that found between students of different schools (Hox, 1995). We consequently rely on a hierarchical linear model (HLM, hereafter) for taking into account the nested structure of the database. In our analysis, data is structured into two levels: students (level 1) and centres (level 2).

$$Y_{ij} = \beta_{0j} + \sum_{k=1}^{n} \beta_{1j} X_{kij} + \varepsilon_{ij} \qquad \varepsilon_{ij} \sim N(0, \sigma^2) \quad (1)$$

$$\beta_{0j} = \gamma_{00} + \sum_{1} \gamma_{01} Z_{1j} + \mu_{0j} \quad \mu_{0j} \sim N(0, \tau_0) \quad (2)$$

$$\beta_{1j} = \gamma_{10} \quad (3)$$

$$Y_{ij} = \gamma_{00} + \gamma_{10} X_{kij} + \gamma_{01} Z_{1j} + \mu_{0j} + \varepsilon_{ij} \quad (4)$$

Y_{ij} is the change in the score in reading skills of student "i" enrolled in school "j" between ages 16 and 10. X_{kij} is a vector of "k" independent variables of the individual level and Z_j is a vector of "l" variables of the school level. Equation 4 is obtained by substituting equations 2 and 3 (level 2) for the β in equation 1 (level 1). In our model specification, we estimate fixed effects (eq. 3).

The dependent variable is the difference between the scores in the reading skills of students between ages 10 and 16. This difference is calculated using the sets of plausible values –random values calculated from the distribution of the results in the assessments- provided by PIRLS and PISA. This gives a total of 25 combinations. Results in PIRLS and PISA are originally scaled to a mean of 500 and standard deviation of 100 *within* each of the respective surveys. A score of 500 points in PIRLS is not equivalent, however, to a score of 500 points in the PISA scale, given the different number of countries participating in the assessments. Therefore, following Brown et al (2007), we tackle this issue using international z-scores for the countries participating in both assessments, that is, normalizing scores for each survey at the student level, with a mean of 0 and a standard deviation of 1 across 25 jurisdictions.

A set of individual, household and school level variables have been included in the model (table A1 in the Appendix). Additionally, interactions between the predicted score in PIRLS 2006 –quartiles- and grade retention have been introduced. This allows us to take into account previous performance and to assess different impacts of grade retention to students with different profiles. This allows us to overcome the reverse causality issue that, as far as we know, affects all the previous literature on the impact of grade retention for Spain. We however acknowledge that, although an effort has been made for introducing into the analysis a wide range of controls, we cannot discard the possibility that unobservable variables may be affecting the results. Individual and school level weights have been applied throughout.

3 Results

This section is structured in two subsections. Subsection 3.1 provides a brief overview of the main characteristics of repeaters and non-repeaters. The results of the hierarchical linear model explained in subsection 2.3 are presented in subsection 3.2.

3.1 Descriptive analysis

A description of the variables used in the analysis may be found at Table A.1 in the appendix. Table 2 provides information on different features in which repeaters and non-repeaters differ. The latter seem to perform worse than non-repeaters at age 10; however the gap between both groups increases greatly between ages 10 and 16. This may be giving an idea of a negative impact of grade retention although we acknowledge, as it may be seen at Table A.1, a reduction in the dispersion of scores at age 10, due to the imputation method applied.

Table 2. Main characteristics of repeaters and non-repeaters

	Non-repeater	Repeater		Non-repeater	Repeater
Mean PIRLS z_score	0.346	0.309	% Did not attend ISCED0	4.32%	7.16%
Mean PISA z_score	0.258	-0.495	% Non-nuclear household	8.28%	14.85%
% Girls	52.13%	45.97%	% Immigrant background	8.55%	16.89%
% Born January to March	25.44%	27.71%	Mean ESCS (index)	0.50	-0.63
% Born October to Dec.	24.59%	25.09%	% Public school	61.31%	77.76%

Source: Self-elaborated from PISA-2012 data.

A non-surprising finding for the Spanish case (see Section 1), it's the fact that boys seem to repeat more frequently than girls. Although not shown in the table, Spanish girls outperform boys in the PISA-2012 reading competence too. Additionally, it is worth mentioning small differences were found in the distribution of repeaters and non-repeaters by date of birth. Some differences were identified, however, when analyzing the percentage of students who did not attend pre-primary education. Nevertheless, the fact that pre-primary education in Spain is, -although not compulsory- nearly universal, reduces the potential weight of this variable for explaining differences between both groups.

The household characteristics seem to be very relevant for understanding grade retention processes. First, the percentage of non-repeaters who live with their both parents is considerably larger than for repeaters. This is in line with studies such as Fernández Enguita et al. (2010). Second, being first or second generation immigrant seems to increase the chances of repeating a course during lower secondary school. Cordero et al. (2014) find differences in the probability of grade retention between first and second generation immigrants. And third, repeaters tend to live in households with a lower socio-economic background. This last statement can be related to the fact that the percentage of repeaters enrolled in public schools is larger than that found at public schools. A number of studies such as Escardíbul and Villarroya (2009) or Mancebón et al. (2010) have described the existence of socio-economic status based student selection processes in Spain not only by private independent schools, but also by the publicly-subsidized.

3.2 Results: HLM

There has been a heated debate in the economics of education literature on the role of grade retention policies in determining student's educational attainment. In this paper, we contribute to this debate by exploiting the rich information in PIRLS-2006 and PISA-2012 studies.

Our empirical approach attempts to quantify the relative importance of grade repetition on academic achievement. In particular, our main outcome of interest is the difference in attainment between primary (PIRLS) and secondary school (PISA). The impact of grade repetition variables are modelled as a HLM that includes a set of individual and household characteristics as well as school characteristics highlighted in previous literature. The complete list of variables used for the empirical analysis together with their description and proportions are given in Table A.1.

Coefficient estimates are presented in Table A.2. Accordingly to our research question, we first focus on the coefficients associated to the grade retention variables. Results clearly show a negative impact of repetition on achievement. That is, for a student who has repeated once during lower secondary education the difference in attainment between primary and secondary education declines 0.36 standard deviations with respect to non-repeaters. Besides, the magnitude of this negative effect increases by more than 50% when the student repeats twice or more years at this educational stage, indicating that the negative effect of grade retention is also cumulative. Furthermore, we have included in our estimation equation interaction terms measuring the impact of repetition along the distribution of scores in primary education (i.e. quartiles). Results indicate that the magnitude of the negative effect is decreasing in prior academic performance affecting more severely the best students among the low achievers. We interpret this finding as a strong negative impact on the motivation and self-esteem of individuals who are retained (Holmes, 1989).

As for the set of remaining individual and household controls, it is worth noting the positive sign of coefficients associated to students living in a household with a “non-standard” family structure (i.e. single parent household), and first generation immigrant households. We believe that the origin of these results comes from our outcome variable, as we are not measuring attainment per se but the difference in academic attainment between primary and secondary school. Then, a positive coefficient associated to students that usually perform below the average (reference) does not mean these students are at the same learning level as their peers in secondary education. Finally, grade repetition seems to affect more to those students coming from a lower socio-economic background.

Regarding school variables, grade retention shows a greater effect on students who live in small cities. Students who attend schools with ICT access and with more autonomy in the allocation of their resources also seem to improve their results between ages 10 and 16.

4 Discussion

Grade retention is widely used in Spain although a) educational acts have considered it a last resource policy and b) there is a lack of consistent studies which assesses its usefulness for improving academic performance and, subsequently, reducing school failure and early school dropout. While social and teachers beliefs and tradition may have played a role in explaining the former (Arregi et al. 2009), the lack of adequate data and methodological issues such as endogeneity and reverse causation justify the latter.

Our results confirm that, once previous performance is taken into account, the sign of the effect of grade repetition on the reading skills of Spanish students is still negative. These results are in line with most of the previous literature on the impact of grade retention described in Section 2. It is also noteworthy that the impact of grade retention during the lower secondary school level by previous performance is heterogeneous. The reading skills of secondary school repeaters who had a relatively better performance during primary school fell relatively more. This may be related to a larger fall in student motivation. Disgracefully, our data do not allow us to go beyond the formulation of this hypothesis. Nevertheless, independently of the mechanism which explains this result, it has a clear policy implication: while teachers should be careful taking decisions on grade retention, they should be especially cautious when making repeat students whose previous achievement was relatively better.

The negative impact of grade repetition and the importance of previous achievement on its effect have important policy implications. The most important of them is probably the need to identify students at risk of grade retention early at the initial stages of the educational system. Results in 3.1 provide some guides on the characteristics of students at risk of grade retention.

Among other possible measures, the identification of students at risk at early stages of the educational system and the introduction of targeted supports and services to students who are automatically promoted may be an alternative to grade retention (Darling-Hammond, 1998).

It is also worth mentioning the stark differences in repetition rates among the Spanish regions. This heterogeneity does not seem to be directly linked to their mean academic performance. There are therefore different "cultures of grade retention" within Spain. Indeed, repetition being finally decided by teachers, it cannot be discarded that grade repetition decisions vary by school, even within the same region –i.e. two students with a similar level of skills, may repeat or not, depending on the centre where they are enrolled. One of the main arguments in favour of applying grade repetition being equality of treatment, this is clearly a hypothesis that will deserve larger attention in the future.

Finally, we acknowledge some limitations of our study. First, we focus on the short-term impact of repetition at the secondary school level. Empirical evidence available for other countries seems to describe different effects of grade retention at earlier stages of the educational system. Grade retention may also have longer-term effects, such as on the probability of getting access into higher education (Andrew, 2014). Second, we analyze reading skills; we cannot discard the effect of grade retention may be heterogeneous by competencies. Third, the validity of results is conditioned by the quality of the imputation method and the comparability of PIRLS and PISA results. Additionally, our study does not control for unobservable variables such as motivation. Nevertheless, while waiting for improved databases, this paper uses an innovative methodology in order to provide solid evidence on the ineffectiveness of grade repetition in Spain and makes a case for reconsidering the use of this policy.

References

Acock, A. (2005). Working with Missing Values. *J. Marr. Fam.*, 67, 1012-1028.

Allen, C. S., Chen, Q., Willson, Hughes V. L. (2009). Quality of Design Moderates Effects of Grade Retention on Achievement: A Meta-Analytic, Multi-Level Analysis. Educational Evaluation and Policy Analysis 31:480–99.

Andrew, M. (2014). The Scarring Effects of Primary-Grade Retention? A Study of Cumulative Advantage in Educational Career, *Social Forces*, 93 (2), 653-685.

Arregi, A., Martínez, P., Sainz, A. and Ugarriza, J. R. (2009). *Efecto de las repeticiones de curso en el proceso de enseñanza-aprendizaje del alumnado*. Bilbao, ISEI-IVEI.

Brown, Giorgina, Micklewright, John, Schnepf, Sylke V., and Robert Waldmann. 2007. "International Surveys of Educational Achievement: How Robust are the Findings?" *Journal of the Royal Statistical Society Series A*, 170 (3), 623–646.

Calero J., Choi A., Waisgrais S. (2010). Determinantes del Riesgo de Fracaso Escolar en España: Una Aproximación a través de un Análisis Logístico Multinivel aplicado a PISA-2006. Revista de Educación, 2010, 225-256.

Calero J., Escardibul J. (2007). Evaluación de Servicios Educativos: el rendimiento en los centros públicos y privados medido en PISA-2003. Hacienda Pública Española, 183, 33-66.

Choi de Mendizábal, A., Calero J. (2013). Determinantes del riesgo de fracaso escolar en España en PISA-2009 y propuestas de reforma. Revista de Educación, 362. Septiembre-diciembre 2013, pp. 562-593.

Cordero, J.M., Crespo, J.M., Santín, D. (2010). Factors Affecting Educational Attainment: Evidence From Spanish Pisa 2006 and Sectoral Economic Studies, Euro-American Association of Economic Development, vol. 10(3).

Cordero, J.M., Manchón, C., and Simancas, R. (2014). La repetición de curso y sus factores condicionantes en España. *Revista de Educación*, 365, 12-37.

Crothers, L.M., Schreiber, J.B., Schmitt, A.J., Bell, G.R., Blasik, J., Comstock, L.A. et al. (2010). A Preliminary Study of Bully and Victim Behavior in Old-for-Grade Students: Another Potential Hidden Cost of Grade Retention or Delayed School Entry. *Journal of Applied School Psychology*, *26*(4), 327-338.

Darling-Hammond, L. (1998). Alternatives to grade retention. *School Administrator*, *55*(7), 18-21.

Dong, Y. (2009). Kept back to get ahead? Kindergarten retention and academic performance. Eur. Econ. Rev. 54 (2), 219–236.

Dupriez, V., Dumay, X. and Vause, A. (2008). How Do School Systems Manage Pupils' Heterogeneity? *Comparative Education Review*, *52*(2), 245-273.

Eide, E., Showalter, M. (2001). The effect of grade retention on educational and labor market outcomes. Econ. Educ. Rev. 20 (6), 563–576.

Escardíbul, J-O, and Villarroya, A. (2009). The inequalities in school choice in Spain in accordance to PISA data. *Journal of Education Policy*, 6, 673-696.

Fernández Enguita, M; Mena, L., and Riviere, J. (2010). *Fracaso y abandono escolar en España*. Barcelona, Fundació La Caixa.

García-Pérez, J. I., Hidalgo-Hidalgo, M., Robles-Zurita, J. A. (2014). Does grade retention affect students' achievement? Some evidence from Spain. Applied Economics, Taylor & Francis Journals, vol. 46(12), pages 1373-1392, April.

Guio Jaimes, J.M., Choi de Mendizábal, A. (2014). Evolución del riesgo de fracaso escolar en España durante la década del 2000: Análisis de los resultados de PISA con un modelo logístico de dos niveles. Estudios sobre Educación, 26, 33-62.

Glick, P., Sahn, D. (2010). Early academic performance, grade repetition, and school attainment in Senegal: a panel data analysis. World Bank Econ. Rev. 24 (1), 93–120.

Gomes-Neto, J.B., Hanushek, E.A. (1994). Causes and consequences of grade retention: evidence from Brazil. Econ. Dev. Cult. Chang. 43 (1), 117–148.

Goos, M., Schreier, B., Knipprath, H., De Fraine, B., Van Damme, J. and Trautwein, U. (2013). Can cross-country differences in the practice of grade retention be explained? A closer look at national educational policy factors. *Comparative Education Review*, 57 (1), 54-84.

Holmes, T.C. (1989). Grade level retention effects: a meta-analysis of research studies. In: Shepard, Lorrie A., Smith, Mary L. (Eds.), Flunking Grades: Research and Policies on Retention. Falmer Press, London, pp. 16–33.

Hox, J. (1995). *Applied Multilevel Analysis.* Amsterdam, TT-Publikaties.

Jacob, B., Lefgren, L. (2004). Remedial education and student achievement: a regression–discontinuity analysis. Rev. Econ. Stat. 86 (1), 226–244.

Jacob, B.A., and Lefgren, L. (2009). The effect of grade retention on high school completion. *Applied Economics*, *1*(3), 33-58.

Jerrim, J., Choi, Á, and Simancas, R. (2014). Two-Sample Two-Stage Least Squares (TSTSLS) Estimates of Earnings Mobility: How Consistent are They? *DoQSS Working Papers*, 17.

Jimerson, S.R., Anderson, G.E. and Whipple, A.D. (2002). Winning the battle and losing the war: examining the relation between grade retention and dropping out of high school. *Psychology in the Schools*, *39*(4), 441-457.

Manacorda, M. (2012). The cost of grade retention. *The Review of Economics and Statistics*, 94 (2), 596-606.

Mancebón, M.J., Calero, J., Choi, Á., Ximénez-de-Embún, D. P. (2012) The Efficiency of Public and Publicly Subsidized High Schools in Spain: Evidence from Pisa-2006. Journal of the Operational Research Society, Vol. 63, Issue 11, pp. 1516-1533, 2012.

Martin, A.J. (2011). Holding back and holding behind: grade retention and students' non-academic and academic outcomes. *Educational Research Journal*, *37*(5), 739-763.

MECD. (2014). *Las cifras de la educación en España 2014*. Madrid, MECD.

Mullis, I., Martin, M., Kennedy, A and Pierre Foy (2007). *PIRLS 2006 International Report*. Chestnut Hill: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

OCDE (2011). When students repeat grades or are transferred out of school: What does it mean to education systems? *PISA in Focus*, 2011/6.

OECD (2014). *PISA 2012 Results in Focus: What 15-year-olds know and what they can do with what they know*. Paris, OECD.

Ou, S.R. and Reynolds, A.J. (2010). Grade Retention, Postsecondary Education, and Public Aid Receipt. *Educational Evaluation and Policy Analysis, 32*(1), 118-139.

Royston, P. (2005). Multiple imputation of missing values: update. *Stata J.,* 5, 188-201.

Royston, P. and White, I. (2011). Multiple Imputation by Chained Equations (MICE): Implementation in STATA. *J. Stat. Soft.*, 45, 1-20.

Rubin, D. (1976). Inference and missing data, *Biometrika*, 63, 581-592.

Rubin, D. (1996). Multiple Imputation After 18+ Years. *J. Amer. Stat. Assoc.*, 91, 473-489.

Salinas, J., Santín D. (2012). Selección escolar y efectos de la inmigración sobre los resultados académicos españoles en pisa 2006. Revista de Educación, 358. Mayo-agosto 2012, pp. 382-405.

StataCorp. (2013). *Stata Multiple Imputation: Reference Manual. Release 13*. Texas: Stata Press.

Schafer, J. (1999). Multiple imputation: a primer. *Stat. Methods Med. Res.*, 8, 3-15.

Sterne, J., White, I., Carlin, J., Spratt, M., Royston, P., Kenward, M., Wood, A. & Carpenter, J. (2009). Multiple imputation for missing data in epidemiological and clinical research: Potential and Pitfalls. *British Med. J.*, 338, b2, 393.

Van Buuren, S., Boshuizen, H. & Knook, D. (1999). Multiple imputation of missing blood pressure covariates in survival analysis. *Stat. Med.*, 18, 681-694.

White, I., Royston, P. and Wood, A. (2011). Multiple imputation using chained equations: Issues and guidance for practice, *Stat. Med.*, 30, 377-399.

Appendix

Table A.1. Summary statistics

	Mean	S.d.	Min.	Max	N
Imputed PIRLS-2006 reading z_scores plausible value 1	0.374	0.005	-0.991	1.076	20,437
Imputed PIRLS-2006 reading z_scores plausible value 2	0.289	0.005	-1.016	0.941	20,437
Imputed PIRLS-2006 reading z_scores plausible value 3	0.312	0.005	-1.055	0.989	20,437
Imputed PIRLS-2006 reading z_scores plausible value 4	0.323	0.005	-0.908	0.972	20,437
Imputed PIRLS-2006 reading z_scores plausible value 5	0.401	0.005	-0.870	1.058	20,437
PISA reading z_scores plausible value 1	0.144	0.020	-3.856	3.220	20,437
PISA reading z_scores plausible value 2	0.142	0.021	-3.733	3.038	20,437
PISA reading z_scores plausible value 3	0.143	0.020	-3.800	3.267	20,437
PISA reading z_scores plausible value 4	0.144	0.021	-3.972	3.121	20,437
PISA reading z_scores plausible value 5	0.140	0.020	-4.233	2.969	20,437
A. Individual and household characteristics					
Did not repeat during lower secondary education	0.847	0.360	0	1	20,306
Repeated one year during lower secondary education	0.136	0.343	0	1	20,306
Repeated two years during lower secondary education	0.172	0.130	0	1	20,306
Sex (girl)	0.512	0.006	0	1	20,437
Born between January and March	0.241	0.005	0	1	20,437
Born between April and September	0.495	0.006	0	1	20,437
Born between October and December	0.246	0.005	0	1	20,437
Did not attend ISCED0	0,048	0,003	0	1	20,285
Attended ISCED0 one year	0,069	0,003	0	1	20,285
Attended ISCED0 more than one year	0,883	0,005	0	1	20,285
Single parent or other situation	0.093	0.003	0	1	19,570
Non-immigrant household	0.914	0.280	0	1	20,234
Immigrant household: first generation	0.074	0.261	0	1	20,234
Immigrant household: second generation	0.012	0.108	0	1	20,234
Language at home: language of the test	0.813	0.016	0	1	20,102
Language at home: international language	0.187	0.016	0	1	20,102
ESCS (Socio-economic status index)	-0.054	0.030	-3.92	2.73	20,437
HEDRES (Home educational resources index)	0.122	0.014	-3.93	1.12	20,437

Table A.1. Summary statistics (continuation)

B. School characteristics

	Mean	**S.d.**	**Min.**	**Max**	**N**
Publicly-subsidized private school	0.652	0.027	0	1	19,789
Independent private school	0.266	0.025	0	1	19,789
Public school	0.082	0.015	0	1	19,789
School size (number of students)	722,795	23,47	45	4,128	19,350
City size (less than 100,000 inhabitants)	0.261	0.024	0	1	20,243
City size (100,000 to 1,000,000 inhabitants)	0.653	0.026	0	1	20,243
City size (more than 1,000,000 inhabitants)	0.086	0.016	0	1	20,243
Student-teacher ratio	12,623	0.342	1.111	139	19,077
Mean years of education of parents	12,570	1.965	5	16.5	20,437
Percentage of immigrant students >20%	0.144	0.019	0	1	20,437
ICTSCH	-0.162	0.024	-2.804	2.826	19,895
Class size (number of students per class)	25,572	0.282	13	48	17,911
Respcurr	-0.436	0.044	-1.26	1.44	20,311
Respres	-0.394	0.039	-0.80	2.71	20,311
Regional dummy: Rest of the country	0.203	0.030	0	1	20,437
Regional dummy: Andalucía	0.194	0.024	0	1	20,437
Regional dummy: Aragón	0.026	0.004	0	1	20,437
Regional dummy: Asturias	0.020	0.003	0	1	20,437
Regional dummy: Baleares	0.021	0.003	0	1	20,437
Regional dummy: Cantabria	0.012	0.002	0	1	20,437
Regional dummy: Castilla y León	:0.049	0.007	0	1	20,437
Regional dummy: Catalunya	0.169	0.022	0	1	20,437
Regional dummy: Extremadura	0.025	0.004	0	1	20,437
Regional dummy: Galicia	0.050	0.007	0	1	20,437
Regional dummy: La Rioja	0.007	0.001	0	1	20,437
Regional dummy: Madrid	0.132	0.018	0	1	20,437
Regional dummy: Murcia	0.031	0.005	0	1	20,437
Regional dummy: Navarra	0.015	0.002	0	1	20,437
Regional dummy: País Vasco	0.047	0.004	0	1	20,437

Source: Self-elaborated from PISA-2012 and PIRLS-2006 data.

Table A.2. Results from the hierarchical linear model; reading competency, PISA-2012.

	Coefficient	Standard E.
Intercept	-0.599***	0.174
A. Individual and household characteristics		
Repeated one year during lower secondary education	-0.363***	0.057
Repeated two or more years during lower secondary education	-0.553***	0.074
Interaction term: Repeated x first quartile in PIRLS score	-0.271***	0.064
Interaction term: Repeated x second quartile in PIRLS score	-0.194***	0.073
Interaction term: Repeated x third quartile in PIRLS score	-0.136***	0.066
Sex (girl=1)	0.191***	0.245
Born between April and September	0.002	0.019
Born between October and December	-0.018	0.023
Attended ISCED0 one year	0.042	0.080
Attended ISCED0 more than one year	-0.004	0.069
Single parent or other situation	0.086***	0.031
Ref. Two-parents in the household		
Immigrant household: first generation	0.193***	0.051
Immigrant household: second generation	-0.063	0.094
Ref. Non-immigrant household		
International language at home (ref. language of the test)	0.059	0.041
ESCS (Socio-economic status index)	0.224***	0.020
HEDRES (Home educational resources index)	0.008	0.011
B. School characteristics		
Publicly-subsidized private school	-0.049	0.045
Independent private school	-0.031	0.070
(Ref. public school)		
School size (number of students)	-0.000	0.000
City size (100,000 to 1,000,000 inhabitants)	0.088***	0.039
City size (more than 1,000,000 inhabitants)	0.233***	0.065
(Ref. less than 100,000 inhabitants)		
Student-teacher ratio	0.000	0.002
Mean years of education of parents	0.020	0.014
Percentage of immigrant students >20%	0.018	0.058
ICTSCH	-0.040***	0.013
Class size (number of students per class)	0.003	0.002
Respcurr	-0.026	0.028
Respres	0.094***	0.039

Source: Self-elaborated from PISA-2012 and PIRLS-2006 data.

Note: ***, statistically significant at 99%; **, 95%; *, 90%. Regional dummies included.