PREDICTORS AND EFFECTS OF GRADE REPETITION
PREDICTORES Y EFECTOS DE LA REPETICIÓN DE CURSO

ABSTRACT

Empirical evidence mostly describes the negative effects of grade repetition on academic performance. Nevertheless, the international use of this educational policy is still widespread. In this study, drawing on data for the Spanish case, we provide additional evidence on the need to consider different policies to grade repetition in order to enhance the achievement of low-performers. These alternative policies should be based on two main principles: individualized treatment and early intervention. Results have been achieved from the application of a novel methodology which allows to combine microdata from two international educational assessments.

Keywords: Grade retention; academic achievement; reverse causality; PISA; PIRLS.

JEL Codes: I24, I28, H52.

RESUMEN

La literatura apunta la existencia de efectos negativos de la repetición de curso sobre el rendimiento académico de los alumnos. Sin embargo, en el ámbito internacional muchos países siguen aplicando esa política de forma intensiva. En este estudio, apoyándonos en datos del caso español, aportamos evidencia adicional acerca de la conveniencia de sustituir la repetición por medidas alternativas, las cuales, a partir de los resultados alcanzados, deberían basarse en dos principios básicos: tratamiento individualizado e intervención temprana. Los resultados se han obtenido a partir del empleo de una metodología novedosa que permite la combinación de los microdatos procedentes de dos evaluaciones educativas internacionales.
.

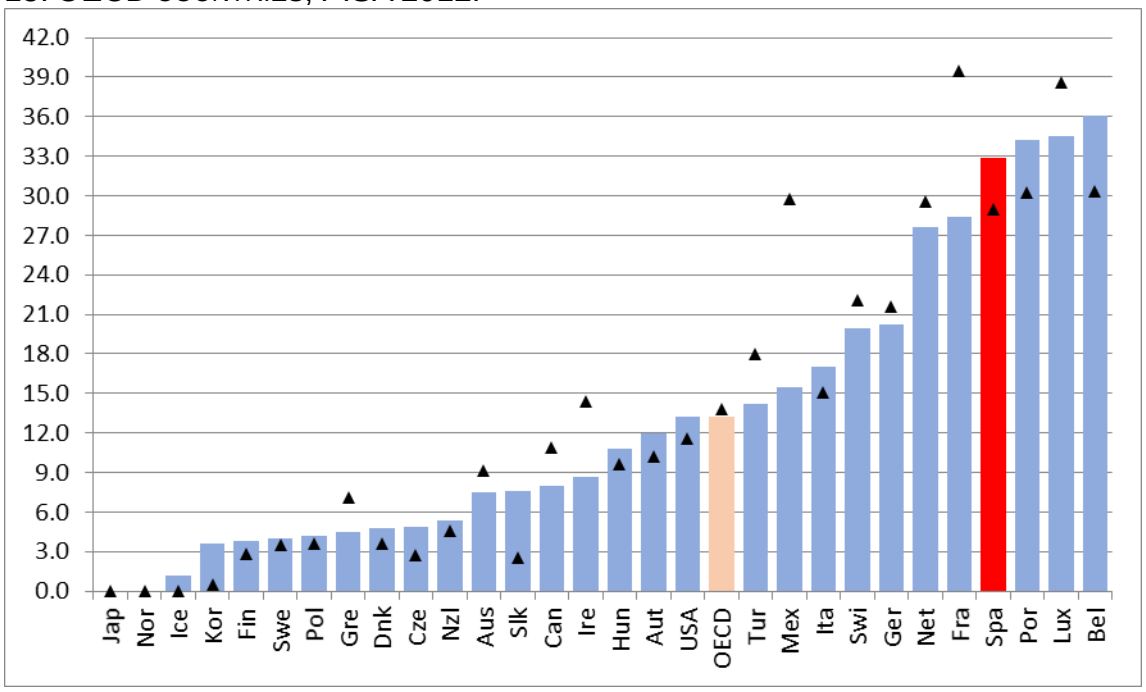
Palabras clave: repetición; rendimiento académico; causalidad inversa; PISA; PIRLS.

Clasificación JEL: I24, I28, H52.

1. INTRODUCTION

The application of grade repetition varies considerably across education systems: while some countries prefer social promotion, others make intensive use of grade retention. Figure 1 illustrates the heterogeneity in the application of grade repetition across countries. While its application is widespread in countries such as Luxembourg, Portugal or Belgium, in other countries such as Japan, Norway or Iceland the use of this policy is scarce.

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.

This variation has been explained in terms of tradition, social beliefs and cultural factors (Goos et al., 2013); yet, the continuing use of grade repetition is striking given that most of the existing literature would seem to highlight its ineffectiveness for increasing student performance (Manacorda, 2012). Indeed, empirical evidence of the effectiveness of grade retention for some of the countries that apply it most intensively is weak.

In this paper we aim to answer two key questions for policymakers. First: who are the students facing a higher risk of grade repetition? And, second: is grade repetition an effective measure for increasing academic performance? Answering the first question will be especially relevant if grade repetition is proven to be ineffective, as it will provide information on the type of measures that may be considered as alternatives to grade repetition. The answer to the second question will allow us to reflect on the adequacy of maintaining grade repetition as a policy.

Regarding our first research question, studies analysing the determinants associated with the decision to retain students in a certain course have focused on academic and non-academic factors, the former being the most relevant for our study. This literature shows that children showing poor academic achievement at the beginning of their schooling have a higher probability of repeating a grade during

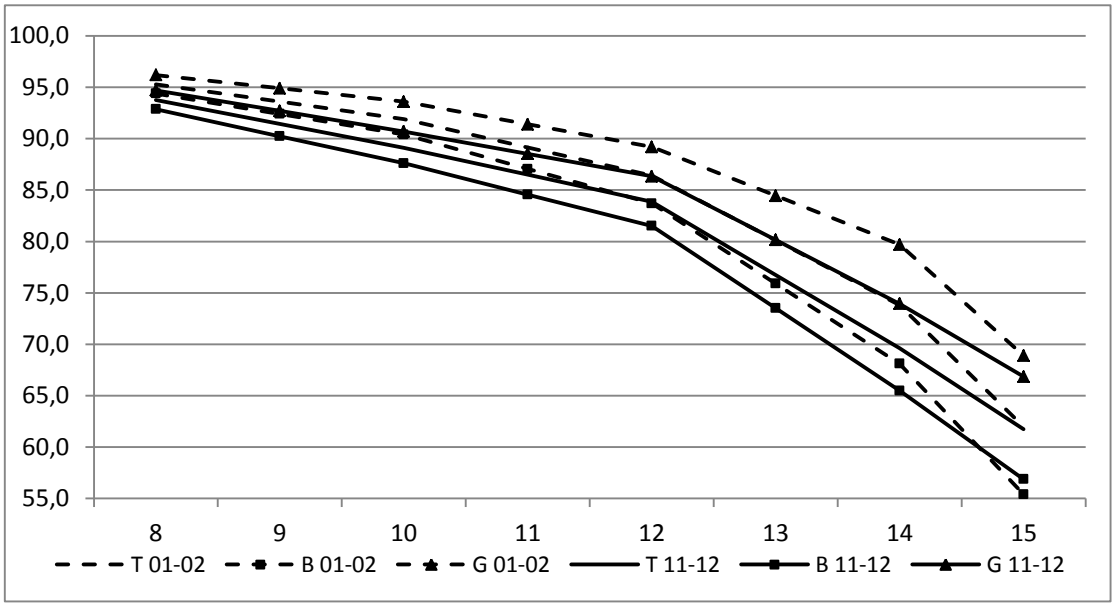
subsequent years in the US (Bali et al., 2005; Ferguson et al., 2001; Frey, 2005; Wilson and Hughes, 2009), China (Chen et al., 2010), South Africa (Liddell et al., 2001) and Brazil (Gomes Neto and Hanushek, 1994).

On the other hand, there is a vast international literature within the economics of education concerned with the effects of grade retention policies. Jacob and Lefgren (2004) show that the results from this literature are inconclusive. Besides, there are also important methodological issues related to the compelling role of reverse causality/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 stronger negative association between retention and academic performance (Allen et al., 2009). Moreover, studies focusing on the estimation of the causal effect of grade retention policies on academic achievement and school dropout also reveal mixed empirical findings (Dong, 2009; Eide and Showalter, 2001; Glick and Sahn, 2010; Gomes-Neto and Hanushek, 1994; Jacob and Lefgren, 2004 and 2009, and Manacorda, 2012).

We will answer both research questions by analysing the Spanish case, a country where almost one third of students have repeated at least one year by age 15. The analysis of Spain as a case study can be useful for other countries with a large proportion of repeaters. Furthermore, data restriction issues for the Spanish case are also common in other countries, thus the methodology used this paper can be replicated with similar data available. Spain has a comprehensive education system that employs grade retention as its main policy for levelling student performance (Dupriez et al., 2008). Figure 2 shows that repetition is mainly used here during compulsory secondary education[1] (ages 12 to 16) and that it is more common among boys. This seems to be consistent with the latter's higher early school dropout rates (26% boys vs. 18.5% girls), in line with studies that describe a positive relationship between grade repetition and early dropout (Jacob and Lefgren, 2009; Ou and Reynolds, 2010). Since 2008, early school dropout figures have fallen (due, in the main, to the economic crisis and high youth unemployment rates), but in 2015 rates for Spanish students of both genders were still well above the EU 2020 Strategy benchmark (reaching 15% in the Spanish case).

[1] Education in Spain is compulsory from ages 6 to 16. It comprises six years of primary school and four years of lower secondary education. Although not 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 one grade during primary school, and two grades during lower secondary education. However, teachers are allowed to promote students that have not passed three subjects. Although the LOE is a national law, there are major differences in retention rates between and within the Spanish regions.

FIGURE 2: PERCENTAGE OF STUDENTS IN THE THEORETICAL AGE FOR EACH GRADE. ACADEMIC YEAR: 2001-02 AND 2011-12.



Source: Based on data from MECD (2014). T: total; B: boys; G: girls.

Empirical research on the effect of grade retention in Spain is fairly sparse. Indeed, there have been only a handful of recent empirical studies that attempt to examine this educational policy (Calero et al., 2010; Calero and Escardíbul, 2007; Choi and Calero, 2013; Cordero et al., 2010; García et al., 2014, Guio and Choi, 2014; Mancebón et al., 2012; Salinas and Santín, 2012). Their main findings point to the negative association of grade retention to academic performance, and how this policy increases the likelihood of school dropout. Notwithstanding, it is important to note that all these studies are limited by the presence of reverse causality: they are unable to determine whether grade retention is the direct cause of poor academic achievement/school failure, or if it is the result of the students' prior characteristics that increase their probability of failure. Therefore, this review leads us to conclude that due to this lack of information there is no robust empirical evidence estimating the effect of grade retention on academic achievement for Spain.

In this paper we overcome significant methodological issues in the existing literature for Spain –namely reverse causation- by merging two repeated cross-sectional international assessments. To the best of our knowledge this is the first Spanish study that analyses the determinants of grade retention controlling for previous academic achievement.

Thus, our contribution to the literature is threefold. First, we identify those students with a higher risk of grade repetition. Second, we provide robust estimates of the effect of grade repetition on the reading performance of Spanish compulsory secondary education students by considering for the first time information on prior academic performance. Finally, we suggest a methodology that may be useful for other countries facing with similar data limitations as Spain.

The headline findings of the research indicate that i) grade repetition has a negative effect on academic performance; ii) this effect is heterogeneous by socioeconomic status (SES) and previous performance; iii) low performance at the primary school level and low-SES, among other determinants, increase the risk of

grade repetition. All these results point out the necessity to re-evaluate the effects of grade retention and consider alternative policies along two main guidelines: a) early intervention and b) individualized treatment.

The remainder of the paper proceeds as follows. Section 2 discusses the methodology and data. Section 3 outlines the results of the analysis of the determinants of grade retention. The results for the effects of grade retention on academic performance are presented in Section 4. Finally, Section 5 concludes.

2. METHODOLOGY AND DATA

For the Spanish case, there are not data offering sufficient information to properly assess the causes of grade retention analysis (i.e., longitudinal data). Thus, we are unable to directly observe students' previous achievement and/or past episodes of grade repetition. In order to overcome this limitation, we will create a pseudo-panel that combines micro-data from the Progress in International Reading Literacy Study (PIRLS) and the Programme for International Student Assessment (PISA). These two databases and the created pseudo-panel are described in the first part of this section; sub-section 2.2 refers to the empirical strategy applied for estimating the determinants of grade retention; finally, sub-section 2.3 is dedicated to the methodology and model applied to study the effects of grade retention.

2.1. DATABASES, PSEUDO-PANEL AND VARIABLES

In 2000, and every three years since, PISA has assessed the competencies acquired by 15-year-old students. In 2012, students from a total of sixty-five countries took tests in reading, mathematics and science. Despite the complete nature of this information, it suffers from a drawback: all PISA data records are cross-sectional in nature, with each wave focusing on the 15-year-old students attending a sampled school. This means individual achievement is not tracked over time. For research purposes and econometric estimations, this absence of panel datasets has a number of limitations, including the impossibility of controlling for reverse causality.

In a similar way to PISA, the International Association for the Evaluation of Educational Achievement (IEA) conducts an international assessment of the reading competencies (PIRLS) of fourth graders in primary education. The first assessment was undertaken in 2001 and thereafter once every five years. PIRLS focuses specifically on a range of reading comprehension strategies applied to literary and informational texts. Forty countries participated in the 2006 round.

To overcome some of the methodological drawbacks of using cross-sectional databases, we construct the pseudo-panel from the PIRLS and PISA datasets. To do so, we adopt a parametric approach and proceed as follows. First, we estimate an educational production function using the auxiliary PIRLS database, the independent variables being those individual and household level variables that are also available in the main PISA sample[2]. We then apply the parameters obtained in this regression to the PISA sample and obtain the predicted score that a student in the PISA database would have obtained on the PIRLS test. Thus, we add a further column to the PISA-2012 database: the students' predicted score on PIRLS-2006. This procedure is repeated five times for each plausible value in PIRLS.

[2] The results of this auxiliary regression are available upon request.

The PIRLS-PISA pseudo-panel for the Spanish case has 25,182 final observations. Missing value issues were detected and corrected in the original databases, using multiple imputation techniques. To make the data comparable, 'repeaters' at primary school level (ISCED1) were deleted from the database, because PISA does not report the age at which a student repeated a grade during primary education. In addition, we deleted from PISA-2012 the first generation of immigrant students who reported having arrived in Spain after 2006, given that they could not have participated in PIRLS-2006. This gave us a final database of 21,230 students.

The variables included in the estimated models capture school characteristics, student backgrounds and student achievements (i.e. reading scores), as provided by PISA-2012, plus the students' prior achievements, as provided by PIRLS-2006. The selection of these variables adheres to the classical educational production function, in which the independent variables are classified at student, household and school levels. The definition of these variables and their summary statistics are reported in Table A1 of the Appendix.

Finally, it should be borne in mind that PIRLS and PISA test results are originally scaled to a mean of 500 and a standard deviation of 100 *within* each of the respective surveys. A score of 500 points on PIRLS is not equivalent, however, to a score of 500 points on the PISA scale, given the different number of countries participating in the assessments. Following Brown et al. (2007), we tackle this by using international z-scores for the countries participating in both assessments.

2.2. ESTIMATING THE DETERMINANTS OF GRADE REPETITION

In the PISA sampling design, 35 students are randomly selected from each school, which in turn has been randomly selected from the pool of national schools making up the population. As sampling is conducted in proportion to school size, larger schools are more likely to be selected; however, students at larger schools are less likely to be chosen to complete the test. For this reason, we need to use a hierarchical model that controls for the correlation between students' results within the same school (Hox, 1995). Specifically, we rely on a hierarchical logistic model so as to take into account the nested structure of the database and the dichotomous nature of the dependent variable.

As indicated above, our data are structured at two levels: that of the students (level 1, *i*) and that of the schools (level 2, *j*). We estimate a logit model where p_{ij}=P (Y_{ij}=1) is the probability of a student repeating a grade in compulsory secondary education. The estimated equations are as follows:

$$log\left[\frac{p_{ij}}{(1-p_{ij})}\right] = \beta_{0j} + \sum_{k=1}^{n}\beta_{1j}X_{kij} + \varepsilon_{ij} \qquad \text{Level 1 (individual) (1)}$$

$$\beta_{0j} = \gamma_{00} + \sum_{1}^{j}\gamma_{01}Z_{1j} + \mu_{0j} \qquad \text{Level 2 (school) (2)}$$

$$\beta_{1j} = \gamma_{10} \qquad \text{Level 2 (school) (3)}$$

$$log\left[\frac{p_{ij}}{(1-p_{ij})}\right] = \gamma_{00} + \gamma_{10}X_{kij} + \gamma_{01}Z_{1j} + \mu_{0j} + \varepsilon_{ij} \qquad (4)$$

where $\log[\frac{p_{ij}}{(1-p_{ij})}]$ is the log of the odds of success – that is, repeating at least one grade in compulsory secondary education – of student "i" enrolled at school "j"; X_{ij} is a vector of "k" independent variables at the individual level and Z_j is a vector of "l" variables at the school level. Equation 4 is obtained by substituting equations 2 and 3 for the β in equation 1. In our model specification, we estimate fixed effects (eq. 3).

2.3. MODELLING THE EFFECTS OF GRADE REPETITION

The effects of grade repetition are estimated using a hierarchical linear model (HLM). The dependent variable now is the difference between the students' scores in reading skills when measured at age 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. The model adopts the following specification:

$$Y_{ij} = \beta_{0j} + \sum_{k=1}^{n} \beta_{1j} X_{kij} + \varepsilon_{ij} \qquad \varepsilon_{ij} \sim N(0, \sigma^2) \quad (5)$$

$$\beta_{0j} = \gamma_{00} + \sum_{1}^{j} \gamma_{01} Z_{1j} + \mu_{0j} \quad \mu_{0j} \sim N(0, \tau_0) \quad (6)$$

$$\beta_{1j} = \gamma_{10} \quad (7)$$

$$Y_{ij} = \gamma_{00} + \gamma_{10} X_{kij} + \gamma_{01} Z_{1j} + \mu_{0j} + \varepsilon_{ij} \quad (8)$$

where Y_{ij} is the change in the reading skills score of student "i" enrolled at school "j" between ages 10 and 15/16. X_{kij} is a vector of "k" independent variables at the individual level and Z_j is a vector of "l" variables at the school level. Equation 8 is obtained by substituting equations 6 and 7 for the β in equation 5. In our model specification, we estimate fixed effects (eq. 7).

The set of individual, household and school level variables included in the model is the same as that reported in Table A1 of the Appendix. Additionally, interactions between the predicted score on PIRLS-2006 – quartiles – and grade retention have been introduced. This allows us to take into account previous performance, thus overcoming the reverse causality that affects all previous studies of the impact of grade retention in Spain, and to assess different effects of grade retention on students of different profiles. However, we acknowledge that, while efforts have been made to introduce a wide range of controls into the analysis, we cannot discard the possibility that unobservable variables may affect the results. Individual- and school-level weights have been applied throughout.

3. THE DETERMINANTS OF GRADE REPETITION

3.1. DESCRIPTIVE ANALYSIS

Table 1 provides information for the most relevant set of characteristics for which 'repeaters' and 'non-repeaters' differ. Repeaters seem to perform worse than non-repeaters at age 9/10, and the gap between the two groups increases over time. These larger raw differences might indicate the potentially negative effect of grade retention on achievement, although we acknowledge, as Table A1 shows, that there is a reduction in the dispersion of scores at age 10, due to the linking method applied that depends on a limited set of variables.

TABLE 1: MAIN CHARACTERISTICS OF STUDENTS WHO HAVE REPEATED A GRADE IN COMPULSORY SECONDARY SCHOOL

	Non-repeater	Repeater		Non-repeater	Repeater
Mean PIRLS z_score	0.310	0.288	% Did not attend ISCED0	0.042	0.065
Mean PISA z_score	0.267	-0.513	% Non-nuclear household	0.085	0.141
% Girls	0.520	0.467	% Immigrant background	0.057	0.115
% Born January to March	0.255	0.266	Mean ESCS (index)	0.059	-0.546
% Born October to Dec.	0.246	0.246	% Public school	0.620	0.773

Source: Based on PISA-2012 data.

By gender, we find that boys tend to repeat a grade more frequently than girls. This result is unsurprising given that in most countries girls outperformed boys on the PISA-2012 reading competence test (OECD, 2014). Small differences were also found by date of birth and by the fact of having attended pre-primary school or not.[3] However, the fact that pre-primary education in Spain is, while not compulsory, nearly universal reduces the potential weight of this variable for explaining differences between the two groups.

As for household characteristics, first, we observe that living with both parents reduces considerably the probability of repeating a grade. Second, being a first- or second-generation immigrant, as opposed to native, increases the likelihood of grade retention in secondary education. Finally, a lower SES is associated with grade retention, while repeaters disproportionately attend public schools. These results are closely related, as disadvantaged students in terms of their SES are more likely to attend public schools. Previous studies have described the existence of SES-based student selection processes in Spain not only by private independent schools, but also by publicly subsidised private schools (Mancebón et al., 2012).

Finally, there also seem to be marked differences in retention rates between the Spanish regions (Table A1 in the Appendix), which might point to the existence of different traditions among teachers and school leaders, and societal beliefs –families support to grade repetition- across the country regarding the benefits of grade retention. Similar results have been found in cross-country studies conducted in

[3] Pedraja et al. (2015) find that the age at which a child starts school in Spain is associated with the likelihood of that student repeating grades. Carabaña (2015) and Cordero et al. (2014) show that not attending pre-primary school increases a student's probability of repeating a grade.

Europe and the OECD (Eurydice, 2011; Goos et al., 2013). This clearly runs counter to one of the main arguments in favour of applying grade retention, namely, equality of treatment.

3.2. MULTIVARIATE ANALYSIS

The results of the logistic regressions are shown in Table 2. While coefficients describe the sign of the relationship between the independent and dependent variables, the odds ratios provide information on the magnitude of these relationships. Results are divided into two blocks: individual and household characteristics, on the one hand, and school-level variables, on the other.

TABLE 2. RESULTS FROM THE LOGISTIC MULTILEVEL MODEL; GRADE REPETITION IN COMPULSORY SECONDARY EDUCATION

	Coeff.	S.E.	O-R
Intercept	-0.034	0.413	0.966
A. Individual and household characteristics			
PIRLS score	-0.959***	0.195	0.383
Sex (1=girl)	-0.259***	0.075	0.771
Born between April and September	-0.166**	0.081	0.847
Born between October and December	-0.017	0.090	0.983
Attended ISCED0 one year	0.258	0.169	1.294
Attended ISCED0 more than one year	-0.013	0.154	0.986
Single parent or other situation (ref. Two-parents household)	0.431***	0.087	1.539
Immigrant household: first generation	0.145	0.170	1.156
Immigrant household: second generation	0.207	0.256	1.230
Ref. Non-immigrant household			
Intl. language at home (ref. language of the test)	-0.115	0.105	0.891
ESCS (Socio-economic status index)	-0.639***	0.041	0.527
B. School characteristics			
School ownership: private (Ref. public school)	-0.110	0.143	0.895
City size (100,000 to 1,000,000 inhabitants)	0.154	0.095	1.166
City size (more than 1,000,000 inhabitants)	0.285*	0.156	1.329
(Ref. less than 100,000 inhabitants)			
School size (number of students)	0.000	0.000	1.000
Student-teacher ratio	-0.048***	0.018	0.952
Mean years of education of parents	-0.028	0.034	0.972
Percentage of immigrant students >20%	0.159	0.119	1.172
Index of ICT school availability (ICTSCH)	-0.005	0.042	0.995
Class size (number of students per class)	-0.007	0.007	0.992
Responsibility of school in curriculum and assessments (RESPCUR)	-0.075	0.055	0.927
Responsibility of school in allocating resources (RESPRES)	-0.010	0.072	0.989

Source: Based on PISA-2012 and PIRLS-2006 data.

Note: *** statistically significant at 99%; ** 95%; * 90%. O-R: Odds Ratio. Regional dummies included.

Our results indicate that individual and household characteristics are highly relevant for explaining grade repetition in compulsory secondary education. Indeed, only one school-level variable (student-teacher ratio) was found to be strongly significant. This is in line with most of the literature on the determinants of academic performance drawing on Spanish PISA data (Cordero et al., 2013).

In the case of the individual-level variables, what is striking is the importance of the students' previous performance on the probability of their repeating a grade in lower secondary school. A one standard deviation increase in reading competencies

at ages 9/10 reduces the risk of grade retention by almost 62%. This result suggests that early intervention could have a substantial effect in reducing high school failure rates in Spain.

Gender differences in terms of the risk of grade retention seem to be important even after controlling for reading competencies at primary school level. In this instance, the probability of repeating a grade is much lower for girls.

A student's month of birth is also statistically significant in our regression, a result that is in line with the findings reported by Pedraja et al. (2015). As expected, SES is highly relevant in accounting for grade retention. Fernandez and Rodriguez (2008) reported similar results using PISA-2003 data.

Living in a household with a "non-standard" structure increases the risk of grade repetition during compulsory education by almost 54%. This is noteworthy as it indicates that family structure makes its effects felt through additional channels besides those of socio-economic and educational resources. This finding is in line with the outcomes reported by Carabaña (2013).

Other variables, such as having attended pre-primary school and immigrant status, were not significant. These results differ from those obtained in previous studies for Spain (see, for example, Cordero et al., 2014). This can be attributed to two factors: first, here we introduce a student's previous performance in our analysis and, second, some of the students with the worst academic performance were dropped due to the procedure adopted in linking PIRLS-2006 to PISA-2012.

Finally, we found evidence of the existence of differences across the Spanish regions.[4] This suggests that grade repetition may depend on other factors that vary across regions, such as different education programs or a different prevalence of the culture of grade repetition. Several studies also report the existence of regional differences in retention rates (Carabaña, 2015; García et al., 2014).

4. THE EFFECTS OF GRADE REPETITION ON READING COMPETENCIES

In this section, we quantify the effect of grade retention during secondary education on students' reading competencies. Thus, our outcome variable is the difference in attainment between primary (PIRLS) and secondary school (PISA). Although the educational production function in our analysis includes a set of individual, household and school level variables, we focus our attention specifically on variables aimed at identifying the effect of grade retention. Coefficient estimates for all the variables are presented in Table 3.

[4] Results not presented in Table 2, but available from the authors upon request.

TABLE 3. RESULTS FROM THE HIERARCHICAL LINEAR MODEL; READING COMPETENCY, PISA-2012

	Coeff.	S.E.
Intercept	-0.662***	0.188
A. Individual and household characteristics		
Repeated one year during lower secondary education	-0.384***	0.037
Repeated two or more years during lower secondary education	-0.653***	0.081
Interaction term: Repeated x first quartile in PIRLS score	-0.411***	0.057
Interaction term: Repeated x second quartile in PIRLS score	-0.274***	0.054
Interaction term: Repeated x third quartile in PIRLS score	-0.209***	0.045
Sex (girl=1)	0.212***	0.025
Born between April and September	0.125***	0.027
Born between October and December	-0.005	0.022
Attended ISCED0 one year	0.043	0.087
Attended ISCED0 more than one year	0.003	0.072
Single parent or other situation	0.073**	0.029
Ref. Two-parents in the household		
Immigrant household: first generation	0.191***	0.054
Immigrant household: second generation	-0.022	0.089
Ref. Non-immigrant household		
International language at home (ref. language of the test)	0.057	0.047
ESCS (Socio-economic status index)	0.214***	0.021
B. School characteristics		
Publicly-subsidised private school	-0.043	0.044
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)	0.071**	0.035
City size (more than 1,000,000 inhabitants)	0.226***	0.070
(Ref. less than 100,000 inhabitants)		
Student-teacher ratio	0.006	0.004
Mean years of education of parents	0.014	0.014
Percentage of immigrant students >20%	0.058	0.056
Index of ICT school availability (ICTSCH)	-0.035***	0.013
Class size (number of students per class)	0.004**	0.002
Responsibility of school in curriculum and assessments (RESPCUR)	-0.051*	0.029
Responsibility of school in allocating resources (RESPRES)	0.105**	0.044

Variances	Null model	Complete model
Schools (μ_{0j})	0.127	0.066
Individuals (ε_{ij})	0.530	0.414
Total ($\mu_{0j} + \varepsilon_{ij}$)	0.657	0.480
% of the total variance explained by the variables		26.90%
% of the level 1 (students) variance explained by the variables		21,88%
% of the level 2 (schools) variance explained by the variables		48.03%

Source: Based on PISA-2012 and PIRLS-2006 data.
Note: *** statistically significant at 99%; ** 95%; * 90%. Regional dummies included.

Our results show a stark negative effect of grade retention on the reading competencies of students. The difference in reading competencies between primary and secondary education of the students who were retained one grade during secondary school is, on average, 0.38 standard deviation points in favour of the non-repeaters. The magnitude of this negative effect increases by more than 65% when a student is retained for two or more grades at this educational stage, indicating that

the negative effect is also cumulative. Furthermore, we account for the possibility that these effects are heterogeneous by including in our estimation equation interaction terms which measure the effect of retention across the distribution of scores in primary education. Our results confirm the presence of heterogeneity indicating that the magnitude of the negative effect is decreasing in prior academic performance, affecting more severely the best students among the low achievers. This suggests that schools should not apply this retention policy homogeneously, especially among their low performing students. We interpret this finding to be an indication of a strong negative effect on the motivation and self-esteem of the individuals that are retained (Holmes, 1989).

To check the robustness of our results, we replicate the analysis by matching PISA-2012 scores with data from the 2007 *Evaluación General de Diagnóstico* (EGD), a national scale assessment tool measuring the performance of Spanish students at age 12. The results obtained in this auxiliary analysis scaffold our earlier findings and are reported in Appendix 2.

As for the remaining individual and household controls, students who live in a non-nuclear family structure and first-generation immigrants show a positive coefficient. Given the definition of our outcome variable, a positive coefficient indicates that the difference in performance of the two respective groups increased slightly. This does not mean these subgroups (students of non-nuclear families and first-generation immigrants) perform better than their reference groups, but that they have closed slightly the gap between ages 9/10 and 15/16. Our measure of socioeconomic background is also statistically significant. This means the SES-based gap in reading competence increases between primary and secondary school. This result is in line with Choi and Jerrim (2016) and stresses the need for the early identification of low SES students as students at risk and, hence, the need to bolster targeted support mechanisms.

5. Discussion and conclusions

Grade retention is widely used in Spain despite the fact that a) educational legislation considers it a policy of last resort and b) there is a lack of consistent studies determining its efficacy for improving academic performance and, subsequently, reducing school failure and early school dropout. While social beliefs and teacher attitudes may play a role in explaining the former (Arregi et al., 2009), the lack of adequate data for addressing methodological issues inevitably account for the latter.

In line with most of the previous literature, our results confirm that, once previous performance is taken into account, the effects of this policy on the reading skills of Spanish students remains negative. This negative effect of grade retention and the importance of previous achievement for this effect have important policy implications. Our results stress the need to reconsider the use of grade repetition as the main policy for levelling students. Not only is it an ineffective policy, it is also unjust, as it has a discriminatory effect by SES. Identifying policies to replace grade repetition falls outside the scope of this paper; however, analysing the educational systems of high performing countries in which retention is seldom applied should provide valuable information on the type of measures that might be considered. Our results may be of interest for countries where the use of grade repetition is widespread, such as Belgium, Portugal or the Netherlands, and stresses the need to consider alternative policies, as social promotion does not seem to be an effective

alternative (Darling-Hammond, 1998). Examples of alternative policies for enhancing the academic achievement of low-performing students are improving teacher practices, such as introducing looping and multi-age grouping (Leuven and Ronning, 2016; Franz et al., 2010); increasing instructional time at school (Slavin et al., 2011; Ritter et al., 2009), out of school (Jacob and Legfren, 2004) or during summer breaks (Borman and Dowling, 2006) for low-performers through individual or group tutoring methods; or making curriculums and educational systems more flexible, removing dead ends and allowing low-performers to have additional time for catching-up with their peers. As it may be seen, the implementation of most of the previous policies requires teachers to identify early students at risk (Allensworth and Easton, 2007; Balfanz et al., 2009).

Indeed, our results highlight the need to identify students at risk of grade retention during the initial stages of their schooling. Students performing poorly at age 9/10 are at greater risk of grade repetition. Moreover, we show that students with a relatively high performance during primary school were among the most negatively affected by grade repetition. This suggests that teachers should exercise extreme caution when deciding which students should be held back a year, especially in the case of students whose prior achievement was relatively strong. This greater fall off in student academic performance may be related to a strong negative effect on the repeaters' motivation. Unfortunately, our data do not allow us to go beyond the formulation of this hypothesis. Finally, it should be borne in mind that most of the worst performing students – those that had already repeated a grade in primary school – were removed from our datasets, which means our estimates represent the lower bound of the negative effect of grade retention. Given these negative effects, grade repetition should be seen as a measure of last resort.

At the same time, our results identify the need for policymakers to increase their efforts to ensure the design of alternative, individualised measures. Our analysis of the determinants of grade repetition indicates that the decision depends on a set of characteristics that extends beyond a student's previous poor academic performance. Students with a low SES are among those at greatest risk of grade repetition and, having repeated a grade, their performance declines more than the average. This result highlights the pressing need to introduce compensatory measures for students presenting these characteristics. A similar case could be made for students living in single-parent households, although our results do not show a greater negative effect of grade retention on students in this group. Among other possible measures, ensuring individualised targeted supports and services are available for students that are automatically promoted may be an alternative to grade retention (Darling-Hammond, 1998).

Finally, we should acknowledge certain limitations in the study reported here. First, we focus on the short-term effects of repetition at the secondary school level. Empirical evidence available from other countries seems to identify different effects of grade retention at earlier stages in the education system. Grade retention may also have longer-term effects, such as an impact on the probability of accessing higher education (Andrew, 2014). Second, we focus solely on reading skills and cannot, therefore, discard the possibility that 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 we await better databases, this paper reports the use of an innovative methodology – which should be of interest for other countries facing

similar data constraints – to provide strong evidence of the ineffectiveness of grade repetition in Spain and it makes a telling case for the reconsideration of this policy.

REFERENCES

Allen, C.S., Chen, Q., Willson, V.L. and Hughes, J.N. (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.

Allensworth, E. M., and Easton, J. Q. (2007): *What matters for staying on-track and graduating in Chicago public high schools: A close look at course grades, failures, and attendance in the freshman year*, Consortium on Chicago School Research, Chicago, IL.

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.

Balfanz, R., Fox, J. H., Bridgeland, J. M., and McNaught, M. (2009): *Grad Nation: A guidebook to help communities tackle the dropout crisis*, America's Promise Alliance, Washington, DC.

Bali, V., Anagnostopoulos, D. and Roberts, R. (2005): "Toward a Political Explanation of Grade Retention. *Educational Evaluation and Policy Analysis*", 27(2), 133-55.

Borman, G. D., and Dowling, N. M. (2006): "Longitudinal achievement effects of multiyear summer school: Evidence from the Teach Baltimore randomized field trial", *Educational Evaluation and Policy Analysis*, 28(1), 25-48.

Brown, G., Micklewright, J., Schnepf, S.V. and Waldmann, R. 2007. "International Surveys of Educational Achievement: How robust are the findings?", *Journal of the Royal Statistical Society Series A*, 170(3), 623–646.

Carabaña, J. (2013): "Repetición de curso y puntuaciones PISA ¿Cuál causa cuál?"; in*: PISA 2012. Informe español. Volumen II: Análisis secundario*, INEE-MECD, Madrid.

Carabaña, J. (2015): "Repetir hasta 4º de Primaria: determinantes cognitivos y sociales según PIRLS", *Revista de la Asociación de Sociología de la Educación*, 8(1), 7-27.

Calero J., Choi A. and 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. and 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.

Chen, X., Liu, C., Zhang, L., Shi, Y. and Rozelle, S. (2010): "Does taking one step back get you two steps forward? Grade retention and school performance in poor areas in rural China", *International Journal of Educational Development*, 30, 544-59.

Choi, Á. and Calero J. (2013): "Determinantes del riesgo de fracaso escolar en España en PISA 2009 y propuestas de reforma", *Revista de Educación*, 362, 562-593.

Choi, Á. and Jerrim, J. (2016): "The use and (misuse) of PISA in guiding policy reform: evidence from Spain", *Comparative Education*, 52(2), 230-245.

Cordero, J.M., Crespo, E. and Santín, D. (2010): "Factors Affecting Educational Attainment: Evidence from Spanish PISA-2006 Results", *Regional and Sectoral Economic Studies*, 10, 55-76.

Cordero, J.M., Crespo, E. and Pedraja, F. (2013): "Rendimiento educativo y determinantes según PISA: Una revisión de la literatura en España", *Revista de Educación*, 362, 273-97.

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.

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", *European Economic Review*, 54, 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., and Showalter, M. (2001): "The effect of grade retention on educational and labor market outcomes", *Economics of Education Review*, 20, 563–576.

Eurydice (2011): *Grade retention during compulsory education in Europe: Regulation and Statistics,* EACEA, Brussels.

Ferguson, P., Jimerson, S.R. and Dalton, M.J. (2001): "Sorting Out Successful Failures: Exploratory Analyses of Factors Associated with Academic and Behavioral Outcomes of Retained Students", *Psychology in the Schools*, 38(4), 327-41.

Franz, D. P., Thompson, N. L., Fuller, B., Hare, R. D., Miller, N. C., and Walker, J. (2010): "Evaluating mathematics achievement of middle school students in a looping environment", *School Science and Mathematics, 110(6)*, 298-308.

Frey, N. (2005): "Retention, social promotion and academic redshirting: What do we know and need to know?", *Remedial and Special Education*, 26(6), 332-46.

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

Guio, J.M. and Choi, Á. (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. and Sahn, D. (2010): "Early academic performance, grade repetition, and school attainment in Senegal: a panel data analysis", *World Bank Economic Review*, 24, 93–120.

Gomes-Neto, J.B. and Hanushek, E.A. (1994): "Causes and consequences of grade retention: evidence from Brazil", *Economic Development and Cultural Change*, 43, 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, 54-84.

Holmes, T.C. (1989): "Grade level retention effects: a meta-analysis of research studies"; in Shepard, Lorrie A., Smith, Mary L. (1989): *Flunking Grades: Research and Policies on Retention*, Falmer Press: London.

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

Jacob, B.A. and Lefgren, L. (2004): "Remedial education and student achievement: a regression–discontinuity analysis", *The Review of Economics and Statistics*, 86, 226–244.

Jacob, B.A. and Lefgren, L. (2009): “The effect of grade retention on high school completion”, *American Economic Journal: Applied Economics*, 1, 33-58.
Leuven, E. and Ronning, M. (2016): “Classroom grade composition and pupil achievement”, *The Economic Journal*, 126 (593), 1164-1192.
Liddell, C. and Rae, G. (2001): “Predicting Early Grade Retention: A Longitudinal Investigation of Primary School Progress in a Sample of Rural South African Children”, *British Journal of Educational Psychology*, 71, 413–28.
Manacorda, M. (2012): “The cost of grade retention”, *The Review of Economics and Statistics* 94, 596-606.
Mancebón, M.J., Calero, J., Choi, Á. and Ximénez-de-Embún, D. (2012): “The Efficiency of Public and Publicly Subsidized High Schools in Spain: Evidence from PISA 2006”, *Journal of the Operational Research Society*, 63, 1516-1533.
MECD. (2014): *Las cifras de la educación en España 2014*. MECD, Madrid.
OECD (2014): *PISA 2012 Results in Focus: What 15-year-olds know and what they can do with what they know*, OECD, Paris.
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.
Pedraja-Chaparro, F., Santín González, D. and Simancas, R. (2015): “Determinants of grade retention in France and Spain: Does birth month matter?”, *Journal of Policy Modeling*, 37(5), 820-34.
Ritter, G.W., Barnett, J.H., Genny, C.S., and Albin, G.R. (2009): “The Effectiveness of Volunteer Tutoring Programs for Elementary and Middle School Students: A Meta-Analysis”, *Review of Educational Research*, 79 (3), 3-38.
Salinas, J. and Santín D. (2012): “Selección escolar y efectos de la inmigración sobre los resultados académicos españoles en PISA-2006”, *Revista de Educación*, 358, 382-405.
Slavin, R.E., Lake, C., Davis, S., and Madden, N. A. (2011): “Effective Programs for Struggling Readers: A Best-Evidence Synthesis”, *Educational Research Review*, 6(1), 1-26.
Wilson, V. and Hughes, J. (2009): “Who Is Retained in First Grade? A Psychosocial Perspective”, *Elementary School Journal*, 109(3), 251-266.

APPENDIX 1

TABLE A1. SUMMARY STATISTICS

	Mean	S.d.	Min.	Max	N
PIRLS-2006 reading z_scores: PV1	0.331	0.276	-0.965	1.100	21.230
PIRLS-2006 reading z_scores: PV2	0.256	0.266	-1.103	0.957	21.230
PIRLS-2006 reading z_scores: PV3	0.285	0.285	-1.151	1.003	21.230
PIRLS-2006 reading z_scores: PV4	0.289	0.266	-1.003	0.994	21.230
PIRLS-2006 reading z_scores: PV5	0.363	0.262	-0.957	1.064	21.230
PISA-2012 reading z_scores: PV1	0.108	0.797	-3.856	3.220	21.230
PISA-2012 reading z_scores: PV2	0.104	0.803	-3.733	3.038	21.230
PISA-2012 reading z_scores: PV3	0.106	0.802	-3.655	3.267	21.230
PISA-2012 reading z_scores: PV4	0.109	0.804	-3.972	3.121	21.230
PISA-2012 reading z_scores: PV5	0.104	0.801	-4.233	2.969	21.230
A. Individual and household characteristics					
Repeated at least one year during lower secondary education	0.206	0.405	0	1	21.230
Sex (girl=1)	0.509	0.500	0	1	21.230
Born between January and March	0.257	0.437	0	1	21.230
Born between April and September	0.497	0.500	0	1	21.230
Born between October and December	0.245	0.430	0	1	21.230
Did not attend ISCED0	0.047	0.211	0	1	21.230
Attended ISCED0 one year	0.065	0.247	0	1	21.230
Attended ISCED0 more than one year	0.887	0.316	0	1	21.230
Single parent or other situation	0.096	0.295	0	1	21.230
Non-immigrant household	0.929	0.256	0	1	21.230
Immigrant household: first generation	0.056	0.231	0	1	21.230
Immigrant household: second generation	0.013	0.113	0	1	21.230
Language at home: language of the test	0.819	0.384	0	1	21.230
Language at home: international language	0.181	0.385	0	1	21.230
ESCS (Socio-economic status index)	-0.065	1.002	-3.18	2.73	21.230
B. School characteristics					
Private school	0.348	0.477	0	1	21.230
Public school	0.644	0.479	0	1	21.230
School size (index of school size)	719.714	408.333	-753.211	4.128	21.230
City size (less than 100,000 inhabitants)	0.264	0.441	0	1	21.230
City size (100,000 to 1,000,000 inhabitants)	0.649	0.478	0	1	21.230
City size (more than 1,000,000 inhabitants)	0.084	0.277	0	1	21.230
Student-teacher ratio	12.289	4.104	1.111	46.933	21.230
Mean years of education of parents	12.542	1.722	5	16.5	21.230
Percentage of immigrant students >20%	0.139	0.347	0	1	21.230
Index of ICT school availability (ICTSCH)	-0.159	0.905	-2.804	3.010	21.230
Class size (students per class)	25.606	5.264	13	48	21.230
Responsibility of school in curriculum and assessments	-0.440	0.728	-1.757	1.440	21.230
Responsibility of school in allocating resources	-0.398	0.631	-1.664	2.710	21.230
Region: Andalucía	0.199	0.400	0	1	21.230
Region: Aragón	0.026	0.158	0	1	21.230
Region: Asturias	0.019	0.139	0	1	21.230
Region: Baleares	0.020	0.142	0	1	21.230
Region: Canarias	0.034	0.182	0	1	21.230
Region: Cantabria	0.011	0.107	0	1	21.230
Region: Castilla-La Mancha	0.051	0.221	0	1	21.230
Region: Castilla y León	0.049	0.217	0	1	21.230
Region: Catalunya	0.163	0.370	0	1	21.230
Region: Comunidad Valenciana	0.115	0.319	0	1	21.230
Region: Extremadura	0.027	0.161	0	1	21.230
Region: Galicia	0.050	0.218	0	1	21.230
Region: Madrid	0.131	0.337	0	1	21.230
Region: Murcia	0.031	0.173	0	1	21.230
Region: Navarra	0.014	0.119	0	1	21.230
Region: País Vasco	0.045	0.208	0	1	21.230
Region: La Rioja	0.007	0.083	0	1	21.230
Region: Ceuta and Melilla	0.004	0.061	0	1	21.230

Source: Based on PISA-2012 and PIRLS-2006.

APPENDIX 2. RESULTS OF THE ANALYSIS COMBINING EGD-2007 AND PISA-2012

We tested the robustness of our results using the *Evaluación General de Diagnóstico* (henceforth EGD), a national student assessment programme of the Spanish education system. By analogy with our main study, we chose the data corresponding to 2007, thus ensuring that we follow a similar cohort to that in our main analysis.

The EGD is conducted with students in the last grade of primary school (12 years old born in 1995) and aims to identify student competencies and knowledge at this educational level in four areas: the Natural, Social and Cultural Environment; Literature and Spanish Language; English Language; and Mathematics. The procedures used in collecting the data and treating the results are similar to those applied in other international assessments.

The sample is obtained by applying a stratified two-stage sampling: in the first stage (private and public) schools are randomly selected within a stratum (in this case the region and the school ownership model); in the second stage, one class is randomly chosen. Then, all students enrolled in this class make up the sample. Once the sample is selected, the students take the standardised tests in each of the four educational areas. They also complete a questionnaire about their attitudes, and personal, social and school background, as do their parents, teachers and school principals. The response rate is above 95% for the target populations.

The three databases differ in a number of aspects. For example, PISA focuses on competencies, PIRLS measures curricular content at the international level, while EDG assesses country specific curricular content. Furthermore, the PISA and PIRLS scores are standardised to z-international scores, whereas the EDG is also standardised but, logically, it does not take into consideration any other countries. Finally, the EDG does not involve the calculation of plausible values in order to measure its outcomes.

The procedures adopted in cleaning and merging the EGD database with the PISA database are the same as those described for PIRLS.

TABLE A.2.1. RESULTS FROM THE HIERARCHICAL LINEAR MODEL; READING COMPETENCY, PISA-2012.

	Coefficient	Standard E.
Intercept	0.641***	0.174
Repeated one year during lower secondary education	-0.220***	0.046
Repeated two or more years during lower secondary education	-0.510***	0.083
Interaction term: Repeated x first quartile in EGD-2007 score	-0.670***	0.070
Interaction term: Repeated x second quartile in EGD-2007 score	-0.521***	0.055
Interaction term: Repeated x third quartile in EGD-2007 score	-0.366***	0.055
Variances	Null model	Complete model
Schools (μ_{0j})	0.187	0.063
Individuals (ε_{ij})	0.697	0.456
Total ($\mu_{0j} + \varepsilon_{ij}$)	0.884	0.519
% of the total variance explained by the variables		41.30%
% of the level 1 (students) variance explained by the variables		34.58%
% of the level 2 (schools) variance explained by the variables		66.30%

Source: Based on PISA-2012 and EGD-2007 data.
Note: *** statistically significant at 99%; ** 95%; * 90%. Regional dummies and usual control variables included.