

# The influence of grade retention on students' competences in Spain

Luis Alejandro Lopez-Agudo<sup>1</sup>  | María Ladrón de Guevara Rodríguez<sup>1,2</sup>  | Oscar David Marcenaro-Gutierrez<sup>1</sup> 

<sup>1</sup>Departamento de Economía Aplicada (Estadística y Econometría). Facultad de Ciencias Económicas y Empresariales, Universidad de Málaga, Málaga, Spain

<sup>2</sup>Programa de Doctorado en Economía y Empresa, Universidad de Málaga, Málaga, Spain

## Correspondence

Luis Alejandro Lopez-Agudo, Departamento de Economía Aplicada (Estadística y Econometría). Facultad de Ciencias Económicas y Empresariales, Universidad de Málaga, Plaza de El Ejido s/n, 29013, Málaga, Spain.

Email: [lopezagudo@uma.es](mailto:lopezagudo@uma.es)

## Funding information

Programa de Doctorado en Economía y Empresa of the Universidad de Malaga; Ministerio de Universidades; Fundación Ramón Areces; Plan Propio de la Universidad de Málaga; European Regional Development Fund; Andalusian Regional Government; Ministerio de Ciencia e Innovación; Universidad de Málaga/CBUA

## Abstract

Grade retention is at the core of the education debate in Spain, to the extent that its impact on students' competences has not been assessed beyond correlation. Because of that, in the present study, we analyse the influence of grade retention on students' competences, using more than 146,000 students from 6 PISA cycles (2003–2018) and an instrumental variable approach, in order to approach a causal influence. Our results show that repeating a grade in Spain seems to reduce students' competences between 1.5 and 1.7 standard deviations. Based on these results, we conclude that the Spanish educational authorities should find an alternative to grade retention, in order to prevent students from attaining a lower competence level due to repetition.

## KEYWORDS

grade retention, instrumental variables, PISA, Spain

## 1 | INTRODUCTION

This research paper aims to analyse how grade retention may influence Spanish students' competences. Traditionally, grade retention has been applied in response to students' underachievement by forcing those students who have not met grade-level proficiency standards to retake the same grade once again instead of moving on to the next one. However, while holding students back a grade is widely used to help them overcome

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Author(s). *European Journal of Education* published by John Wiley & Sons Ltd.

educational difficulties, it depends on the legislation in force in each country. There are countries – for example, the United Kingdom or New Zealand – where grade retention is almost an exception, and others – for example, Denmark or Norway – where it is not allowed. In contrast, grade retention is common practice in France or Portugal, even in elementary school.

In this context, while country-level factors such as education policy are key for grade repetition, students' own characteristics also play a major role. Other factors such as gender (González-Betancor & López-Puig, 2016; Martin, 2009; Tingle et al., 2012), socio-economic status (Mattison et al., 2018; Pedraja-Chaparro et al., 2015) or month of birth have also been found as a driver of grade retention (Jerrim et al., 2022; Ponzo & Scoppa, 2014). This is especially relevant in the case of Spain, as students born in the same calendar year start school in the same academic year. Therefore, as the cut-off is on 1st of January, those students born in January will be almost a year older than their peers born in December. As a consequence, poorer maturity may cause younger students to have greater learning difficulties, which may increase their likelihood of grade repetition (e.g. Bedard & Dhuey, 2006; Calsamiglia & Loviglio, 2019; Cook & Kang, 2016; Sprietsma, 2010).

Particularly in Spain, grade retention has been commonly used. According to the Programme for International Student Assessment (PISA) report (2018), 28.7% of students aged 15–16 had repeated at least one grade. This is almost three times the OECD (i.e. Organisation for Economic Co-operation and Development) average (11.4%) and is only exceeded at the European level by Belgium and Luxembourg (OECD, 2020a). The high rate of students who have been held back has called into question the viability of the Spanish education system and has fuelled an intensive debate. As a result, recent Government educational reforms (published in the Official State Bulletin, or *Boletín Oficial del Estado* in Spanish, i.e. BOE) are likely to alter the Spanish education system in the near future, making grade retention an almost exceptional practice (art. 9 and art. 11; BOE, 2021). However, this new reality does not change the fact that, over the last decade, Spain has been at the top of the European grade retention ranking.

Grade retention has been a way to deal with those students who fail, giving them the opportunity to catch up with their peers and acquire grade-related knowledge. But its benefits are unclear; not only does it significantly increase educational costs but it can also lead to socio-emotional problems among students (Goos et al., 2021). Students who repeat grades may feel like 'academic failures' compared to their peers who pass. As a result, grade retention may negatively affect students' self-esteem (Peixoto et al., 2016; Robles-Piña, 2011), academic self-concept (Lamote et al., 2014; Martin, 2011) or behaviour (Inglés et al., 2015; Jimerson et al., 2005). And, above all, grade retention can affect students' competences.

Most studies report that grade retention's effects on academic achievement are predominantly negative. Chen et al. (2010) found that grade retention did not have a positive effect on students' academic performance in both the short term (the year immediately following the student's retention) and the long term. In fact, the authors found that grade retention (especially in second grade) had a statistically significant and negative effect on student achievement. Likewise, Lamote et al. (2014) analysed the academic performance of 6,411 secondary school students in the Netherlands; the authors found that grade retention did not translate into higher academic performance not only in the short term but also in the long term academic performance actually declined. In the same vein, Taniguchi (2015) found that 5th and 7th grade students in rural Malawi who had repeated a grade had lower academic achievement than those students who had been promoted. Furthermore, the authors identified that absences and being the younger sibling increased the likelihood of repeating 5th grade, while lack of educational resources and parental education were key factors in the likelihood of repeating 7th grade. The same conclusion was reached by Hwang and Cappella (2018); using propensity score methods and a sample of students from kindergarten through 8th grade, they compared similar students and concluded that grade retention had a statistically significant negative effect on reading achievement.

Other studies have found that grade retention has no significant effect on students' academic performance (Allen et al., 2009; Jacob & Lefgren, 2004; Jimerson, 2001; Jimerson & Kaufman, 2003; Jimerson et al., 2005; Wu et al., 2008). Specifically, Jimerson (2001) conducted a meta-analysis of 20 studies, finding that grade retention

did not result in better outcomes than automatic promotion. In the same vein, Allen et al. (2009) conducted a meta-analysis of 22 studies on grade retention, concluding that the association between grade retention and students' academic achievement was null. Similarly, Jacob and Lefgren (2004), using a regression discontinuity design, found no consistent short-term differences in the performance of repeaters in Chicago public schools versus those who were promoted. Eide and Showalter (2001) analysed the effects of grade retention on the probability of dropping out of high school in the United States. For this purpose, they account for the endogeneity of grade retention using an instrumental approach, where the key instrument is the variation across states in kindergarten entry dates; their results showed that grade retention does not seem to have any influence of students' dropout. Regarding the research paper by Klapproth et al. (2016), they used longitudinal data and a sophisticated design of propensity score matching that allowed for carefully matching retained and promoted students in secondary education in Luxembourg. They found that retaining students resulted in short-term benefits for the retained students in terms of school marks, but that the matched promoted students performed equally well in the medium term. The results for the standardised achievement tests also showed that retained students did not differ significantly in their achievement from the students who were promoted, whereas the self-concept of promoted students was higher than that of retained students.

Some studies have found positive effects of grade retention on students' performance (Lorence, 2014; Mahjoub, 2017; Schwerdt et al., 2017). For instance, Schwerdt et al. (2017) used a regression discontinuity design to analyse the causal effect of third-grade retention on students' future outcomes up to 6 years later in Florida public schools. The authors found that, during the first 3 years after being retained, repeater students managed to outperform their age-matched peers who were promoted by 0.23 standard deviations (SD) in mathematics and 0.31 SD in reading. However, these positive effects vanished over time, becoming statistically insignificant in both subjects within 5 years.

## 1.1 | Grade retention and students' competences

In this context, grade retention not only may have influence on students' academic achievement but also on their competences, which are the focus of the present study. According to PISA, competences<sup>1</sup> are defined as: 'students' capacity to apply knowledge and skills in key subjects, and to analyse, reason and communicate effectively as they identify, interpret and solve problems in a variety of situations' (OECD, 2017, p. 13).<sup>2</sup> Therefore, competences differ from content-based knowledge (which is more related to academic achievement) to the extent that the latter is referred the content learned from the school curriculum to solve predefined and structured exercises and problems, whereas competences would refer to the ability to apply this knowledge to new situations and daily life problems.

Focusing on Spain (the country under analysis), although grade retention has been widely applied and increasingly studied, most research studies have not assessed its impact on students' competences beyond correlation. Particularly, Calero et al. (2010) analysed Spanish secondary school students using data from PISA 2006 and found that grade retention significantly explained low competences while promoting dropout. Carabaña (2013) analysed Spanish secondary school students using data collected in PISA 2012. The author concluded that grade retention in primary school was not only associated with cognitive and non-cognitive factors – for example, students' competences or students' age – but was also the best predictor of grade retention in secondary school. Similarly, Gortazar (2019) conducted a descriptive and correlation analysis using PISA 2015 data, finding that there was a positive relationship between grade retention and student dropout, while a negative one between grade retention and students' competences.

In addition, some authors have reached the same conclusion using different methodologies. For instance, García-Pérez et al. (2014) employed a switching regression model to analyse the effect that grade retention could have on the mathematics competences of 15-year-old Spanish students. They found that grade retention

has a negative impact on educational outcomes, with the impact being greater when grade retention was in primary rather than secondary school. Diris (2017) analysed the influence of grade retention on students' competences for 16 countries (including Spain) using PISA 2003–2012 data and the month of birth as instrumental variable. The author found that grade retention in primary school harms students' competences across the distribution, while delayed school entry can produce positive results for those at the lower end of the distribution.

Based in this evidence, it is noticeable that grade retention may have a negative influence – at least in Spain – on students' competences. This only calls into question the efficacy of grade retention for reducing the risk of school failure, especially considering that grade retention can act as a negative trigger, with scarring implications for students' psychosocial development and educational performance through damage in their self-esteem and their school engagement (Andrew, 2014; Goos et al., 2013; Klapproth et al., 2016).

In this context and, given the limited evidence available in Spain, which does not go beyond a correlational analysis, we will try to analyse the actual impact that grade retention has on students' competences, using an instrumental variables methodological approach.

## 1.2 | The contribution of the present study

Specifically, the research question that we want to answer is the following:

- Does grade retention negatively influence 15-year-old students' competences in Spain?

The present study contains some novelties: First, we analyse the influence of grade retention on students' competences focusing on Spain, going beyond correlation and approaching causality using as instrumental variable the month of birth. Second, the data employed come from six PISA waves from the years 2003 to 2018, reaching to more than 146,000 15-year-old students, which enhances the robustness of our results through time.

It has to be highlighted that the novelty of our research study may seem to overlap that of Diris (2017), to the extent that it deals with a similar topic and uses a similar instrumental variable (month of birth), making use of the variability present in 16 countries (including Spain) and reaching a unique conclusion for all the countries regarding the relationship between students' competences and grade retention. Nevertheless, their results could be undermined by some issues: First, the paper relies on the assumption of a homogeneous maturity effect across countries, which may actually differ across countries (as indicated by e.g. Bedard & Dhuey, 2006). Therefore, although the author performed a test which indicates a homogenous effect of maturity across the countries under study, the influence of the month of birth may actually be different for certain countries, especially for Spain, to the extent that children begin school during the calendar year of their 6th birthday and there is no possibility for parents to do otherwise. As reported in 'Exhibit 2: Grade Assessed and Average Age of the Students Assessed in PIRLS 2016' (Mullis et al., 2017), the rest of the countries in the sample of Diris (2017) allow certain flexibility for parents to choose when to enrol their children in compulsory education, potentially influencing this effect of the month of birth. Second, Diris (2017) did not use PISA's recommended practices regarding weighting (i.e. senate weights and balanced repeated replication weights). As highlighted by PISA technical reports (e.g. OECD, 2014), senate weights are needed when performing cross-country regressions with PISA data (Jerrim et al., 2017; OECD, 2020b), while balanced repeated replication weights are needed to calculate correct standard errors with PISA data. Third, our results would be more robust for Spain, as we employ data from 2003, 2006, 2009, 2012, 2015 and 2018 PISA waves (a total of 149,868 observations), whereas the data in Diris (2017) only include 2003, 2009 and 2012 (61,991 observations) for Spain.

The rest of the paper is structured as follows: participants are described, followed by data analysis, procedures employed and results, their discussion and conclusions.

## 2 | METHODS

### 2.1 | Participants

The Programme for International Student Assessment (PISA) is conducted by the OECD and assesses 15-year-old students' competences for the participating countries in reading, mathematics and science. In order to participate, countries are required a sample of at least 42 15-year-old students from each participating school, and a minimum of 150 schools. These students take a cognitive test in the three domains and also answer a student questionnaire. Furthermore, head teachers answer a school background questionnaire.<sup>3</sup> PISA has been conducted in a three-year basis since 2000, reaching to 7 waves in the year 2018. In the present study we employ all the PISA waves in which Spain participated (2003, 2006, 2009, 2012, 2015 and 2018).

The variable in which we base our analysis is that regarding grade retention, which is asked in the student questionnaire. Grade retention education laws did not change in all the PISA cycles under analysis, that is, students can repeat once in primary education<sup>4</sup> and twice in secondary education,<sup>5</sup> being this decision based on an objective criterion: students that failed a minimum of three subjects will have to repeat (or if they failed two subjects, when these are reading and mathematics). Therefore, pooling together six PISA cycles (2003, 2006, 2009, 2012, 2015 and 2018) in which the grade retention laws were kept enhances the precision of our results.

As we are analysing the influence of grade retention on students' scores, we employ the PISA scores provided by the OECD in reading, mathematics and science. These scores were created to have mean 500 and standard deviation 100. In the present study, these scores have been standardised to have mean 0 and standard deviation 1 – using the mean and standard deviation of the population – with the aim of interpreting the results as effect sizes, which is useful for international comparisons. [Table A1](#) (Appendix A) presents the distribution of participants by gender and Spanish region.

### 2.2 | Data analysis

The descriptive statistics for the whole sample, and also by grade retention, can be found in [Table A2](#) (Appendix A). Looking at these statistics, we can see that repeater students present lower scores in all the domains, and that there is a higher proportion of repeater students among boys, from immigrant and lower socio-economic status backgrounds (in terms of parental level of education and number of books at home), who attend public schools and were born in the last months of the year.

### 2.3 | Procedure

#### 2.3.1 | Ordinary least squares

In order to analyse the influence of grade retention on students' competences, the following model is specified:

$$C_{ijrt} = \alpha + \beta GR_{ijrt} + \gamma X_{ijrt} + \delta F_{ijrt} + \vartheta SCH_{ijrt} + \theta PISA_t + \phi REG_{rt} + \varepsilon_{ijrt} \quad (1)$$

where  $i$  is the individual,  $j$  the school,  $r$  the Spanish region and  $t$  the PISA cycle ( $t = 1$  for PISA 2003,  $t = 2$  for PISA 2006,  $t = 3$  for PISA 2009,  $t = 4$  for PISA 2012,  $t = 5$  for PISA 2015 and  $t = 6$  for PISA 2018);  $C_{ijrt}$  are students' standardised competences in reading, mathematics and science (alternatively);  $GR_{ijrt}$  is a dummy variable which takes the value '1' when the student has repeated and '0' otherwise;  $X_{ijrt}$  are students' background characteristics (i.e. sex, immigrant status);  $F_{ijrt}$  are family characteristics (level of education of the father and the mother, number of books at

home);  $SCH_{jrt}$  are school characteristics (school funding);  $PISA_t$  is the PISA cycle;  $REG_{rt}$  is the Spanish region;  $\varepsilon_{ijrt}$  is the idiosyncratic error term.

The estimated  $\beta$  coefficient from this model would show the influence of grade retention on students' competences in reading, mathematics and science, controlling for a wide array of observable characteristics. However, there are potential unobservables included in  $\varepsilon_{ijrt}$  which have been omitted from the model and might bias this  $\beta$  coefficient (e.g. students' ability). This issue can be seen, for instance, if we consider a student who is not mature enough and repeats. In this case, the model in Equation (1) would omit the student's maturity, which is correlated to grade retention; therefore, grade retention would be correlated with the error term, and ordinary least squares (OLS) estimates of  $\beta$  might be biased. This usually happens when working with cross-sectional data, as many authors have indicated (Cordero & Pedraja, 2019; Hanchane & Mostafa, 2010; Lounkaew, 2013). Because of that, an instrumental variable approach has been applied to solve this problem, by the use of two-stage least squares (2SLS).

### 2.3.2 | Two-stage least squares

This methodology requires the identification of an instrument ( $Z_{ijrt}$ ) and the use of other control variables ( $X_{ijrt}$ ,  $F_{ijrt}$ ,  $SCH_{jrt}$ ,  $PISA_t$  and  $REG_{rt}$ ) to remove any potential confounder on the relationship between students' competences and grade retention. The instrument that is going to be used is the month of birth of the student (denoted as  $Z_{ijrt}$ ). This is a categorical variable which is decomposed into binary variables, each one representing a month. In order to be used in the two-stage least squares methodology, this instrument has to fulfil some properties that are described in the following:

- The *relevance condition* or *first stage*. According to this condition, the instrument should be strongly associated with the treatment variable ( $GR_{ijrt}$ ). This happens in the present study, as grade retention is strongly linked to students' month of birth ( $Z_{ijrt}$ ). In particular, this can be seen in Table A3 (Appendix A): Those students who were born later in the year are more likely to repeat. If we compare January with December, there are 10% more repeaters among younger students (December) compared to older students (January).<sup>6</sup> This point is going to be formally illustrated using the Stock and Yogo (2005) test of weak instruments presented in the results section below.
- The *independence/exogeneity assumption*. This condition indicates that the instrument is randomly assigned or 'as good as randomly assigned', which means that it is uncorrelated to the omitted variables we might like to control for. In the present study, month of birth might be considered as good as randomly assigned, to the extent that, in Spain, students are randomly born along the year, that is, parents do not plan the month of birth of their children subject to variables such as their socio-economic status or other characteristics. This can be seen in the first column of Table A2 (Appendix A), in which students' birth months are uniformly distributed along the year.<sup>7</sup>
- The *exclusion restriction*. This condition establishes that there is a sole channel (i.e. through  $GR_{ijrt}$ ) for the influence of the instrument (i.e. month of birth,  $Z_{ijrt}$ ) on the outcome (students' standardised reading, mathematics and science scores,  $C_{ijrt}$ ). This single channel requires that the independence assumption is fulfilled, so that other potential channels for this influence have been controlled for ( $X_{ijrt}$ ,  $F_{ijrt}$ ,  $SCH_{jrt}$ ,  $PISA_t$  and  $REG_{rt}$ ).
- The *monotonicity property* (Barua & Lang, 2016; Dhuey et al., 2019; Fiorini & Stevens, 2014). Barua and Lang (2016) highlighted that 'while the instrument may have no effect on some individuals, all of those who are affected should be affected unidirectionally' (p. 348). This is also called as the *no defiers assumption*, which means that there are no students who decide not to repeat a grade when they are assigned to by their teachers and, likewise, if they are not assigned by their teachers to repeat a grade, they always decide to repeat. This monotonicity property is always fulfilled in our data, as students are not able to choose whether to repeat or not.

After defining this instrument, the model in Equation (1) can then be estimated using two-stage least squares (2SLS). The *first stage* is defined as:

$$GR_{ijrt} = \pi_0 + \pi_1 Z_{ijrt} + \pi_2 X_{ijrt} + \pi_3 F_{ijrt} + \pi_4 SCH_{ijrt} + \pi_5 PISA_t + \pi_6 REG_{rt} + \omega_{ijrt} \quad (2)$$

where  $\omega_{ijrt}$  is the idiosyncratic error term. This equation is estimated and a prediction of the grade retention variable is obtained ( $\widehat{GR}_{ijrt}$ ), which will be included in the model of Equation (1) to create the *reduced form* (hereinafter the base model):

$$C_{ijrt} = \alpha + \beta \widehat{GR}_{ijrt} + \gamma X_{ijrt} + \delta F_{ijrt} + \vartheta SCH_{ijrt} + \theta PISA_t + \varphi REG_{rt} + \varepsilon_{ijrt} \quad (3)$$

where  $\beta$  indicates the influence of grade retention on students' competences in reading, mathematics and science, respectively. These estimations have been performed using OECD recommended practices, that is, final student weights, balanced repeated replication weights and plausible values (OECD, 2020c). For this purpose, Stata software, together with the 'repest' command (which accounts for all these OECD recommended practices) have been employed. The syntax used for the regression models has been displayed in Appendix B.

## 2.4 | Results

Firstly, and following the methodology explained in the previous section, we have estimated our base model (Equation 1) using OLS. As we can observe in Table A4 (Appendix A),<sup>8</sup> grade retention has a negative influence on students' competences between 0.83 and 0.91 standard deviations (SD). Specifically, grade retention reduces performance in mathematics (0.91 SD) more than in science and reading (0.83 SD in both). In order to complement our results, we have distinguishing by gender our main estimations (see Table A5, Appendix A), finding that grade retention negatively influences boys' competences slightly more than girls'. While grade retention worsens boys' performance between 0.85 and 0.93 SD, it reduces girls' performance between 0.80 and 0.88 SD.

However, we have to bear in mind that these estimates are far from being accurate, to the extent that they may suffer from an important variable omission bias. By using OLS, we may have omitted confounders that may be decisive in how grade retention affects students' competences. In this sense, we have estimated our base model (from Equation 3) using two-stage least squares to obtain more accurate estimates.

First, we focus on the first stage results obtained from Equation 2 and presented in the first column of Table 1. As noted in the literature (Bernardi, 2014; Jerrim et al., 2022; Manacorda, 2012; Pedraja-Chaparro et al., 2015), Table 1 shows that older students (i.e. those students born in January) are 9.4% less likely to repeat than those born in December. These results also show that the instrument is relevant to explain grade retention, which supports the relevance condition.

As for the influence that grade retention has on students' competences, we can observe in Table 1 that it negatively influences students' competences but, unlike the results we obtained using OLS, when using two-stage least squares the negative influence that grade retention has on students' competences is even larger. In particular, grade retention is negatively associated with reading performance by 1.5 SD, by 1.68 SD in mathematics and by 1.74 SD in science performance, thus being science competence the most adversely affected (in contrast to OLS estimates). The Stock and Yogo (2005) test of weak instruments also supports that the instrument fulfils the relevance condition.<sup>9</sup>

Regarding the first stage results by sex presented in Table 2, month of birth has a greater influence on the probability of boys repeating than girls. Older male students are 10.6% less likely to repeat than their younger peers, whereas older female students are only 8.4% less likely to repeat than those born at the end of the year. Moreover, in the reduced form results of Table 2 grade retention seems more detrimental for boys' competences

TABLE 1 The influence of grade retention on students' standardised scores, instrumental variables.

Variables	First stage	Reduced form		
		Reading	Mathematics	Science
Repeater (ref.: non-repeater)	-	-1.522*** (0.070)	-1.679*** (0.069)	-1.742*** (0.074)
Month of birth (ref.: December)				
January	-0.094*** (0.005)	-	-	-
February	-0.090*** (0.006)	-	-	-
March	-0.085*** (0.005)	-	-	-
April	-0.079*** (0.005)	-	-	-
May	-0.072*** (0.005)	-	-	-
June	-0.054*** (0.005)	-	-	-
July	-0.050*** (0.005)	-	-	-
August	-0.033*** (0.005)	-	-	-
September	-0.037*** (0.005)	-	-	-
October	-0.035*** (0.005)	-	-	-
November	-0.010* (0.005)	-	-	-
Missing flag. Month of birth	-	-	-	-
Female	-0.084*** (0.002)	0.158*** (0.007)	-0.312*** (0.007)	-0.244*** (0.008)
Missing flag. Female	-0.433 (0.318)	0.313 (0.598)	0.116 (0.589)	0.008 (0.631)
Immigrant status (ref.: native)				
First-generation immigrant	0.148*** (0.005)	0.010 (0.013)	-0.021 (0.013)	-0.000 (0.014)
Second-generation immigrant	0.063*** (0.008)	0.051*** (0.016)	0.003 (0.016)	0.049*** (0.017)
Missing flag. Immigrant status	0.140*** (0.009)	-0.298*** (0.020)	-0.329*** (0.020)	-0.273*** (0.021)
Father level of education (ref.: none)				
ISCED 1	-0.079*** (0.006)	0.018 (0.012)	0.011 (0.012)	-0.015 (0.013)

TABLE 1 (Continued)

Variables	First stage	Reduced form		
		Reading	Mathematics	Science
ISCED 2	-0.089*** (0.006)	0.050*** (0.013)	0.050*** (0.013)	0.018 (0.013)
ISCED 3b, c	-0.121*** (0.008)	0.035** (0.017)	0.049*** (0.017)	-0.003 (0.018)
ISCED 3a, 4	-0.114*** (0.006)	0.061*** (0.014)	0.040*** (0.014)	0.026* (0.015)
ISCED 5b	-0.113*** (0.006)	0.039*** (0.014)	0.019 (0.014)	0.021 (0.015)
ISCED 5a, 6	-0.148*** (0.006)	0.064*** (0.016)	0.089*** (0.015)	0.045*** (0.016)
Missing flag. Father level of education	-0.008 (0.008)	-0.007 (0.015)	0.021 (0.015)	0.001 (0.016)
Mother level of education (ref.: none)				
ISCED 1	-0.057*** (0.007)	0.169*** (0.013)	0.086*** (0.013)	0.144*** (0.014)
ISCED 2	-0.075*** (0.006)	0.175*** (0.013)	0.120*** (0.013)	0.162*** (0.014)
ISCED 3b, c	-0.107*** (0.008)	0.184*** (0.017)	0.107*** (0.017)	0.178*** (0.018)
ISCED 3a, 4	-0.119*** (0.007)	0.169*** (0.015)	0.096*** (0.015)	0.156*** (0.016)
ISCED 5b	-0.104*** (0.007)	0.122*** (0.015)	0.041*** (0.015)	0.129*** (0.016)
ISCED 5a, 6	-0.159*** (0.007)	0.169*** (0.017)	0.117*** (0.017)	0.157*** (0.018)
Missing flag. Mother level of education	-0.001 (0.010)	-0.075*** (0.018)	-0.074*** (0.018)	-0.038* (0.019)
Number of books at home (ref.: 0-10 books)				
11-25 books	-0.103*** (0.005)	0.151*** (0.012)	0.116*** (0.011)	0.086*** (0.012)
26-100 books	-0.227*** (0.004)	0.309*** (0.018)	0.278*** (0.018)	0.230*** (0.019)
101-200 books	-0.293*** (0.005)	0.417*** (0.023)	0.400*** (0.022)	0.350*** (0.024)
201-500 books	-0.331*** (0.005)	0.504*** (0.025)	0.510*** (0.025)	0.450*** (0.027)
More than 500 books	-0.322*** (0.006)	0.495*** (0.025)	0.525*** (0.025)	0.475*** (0.027)

(Continues)

TABLE 1 (Continued)

Variables	First stage	Reduced form		
		Reading	Mathematics	Science
Missing flag. Number of books at home	-0.124*** (0.013)	0.074*** (0.026)	0.092*** (0.025)	0.074*** (0.027)
School funding (ref.: public)				
Private	-0.108*** (0.005)	0.071*** (0.012)	0.025** (0.011)	0.011 (0.012)
Semi-private	-0.086*** (0.003)	0.044*** (0.008)	-0.012 (0.008)	-0.034*** (0.008)
Missing flag. School funding	-0.057*** (0.006)	0.002 (0.011)	-0.025** (0.011)	-0.069*** (0.012)
PISA cycle (ref: PISA 2003)				
PISA 2006	0.094*** (0.005)	0.221*** (0.012)	0.275*** (0.012)	0.248*** (0.012)
PISA 2009	0.055*** (0.005)	0.150*** (0.011)	0.193*** (0.010)	0.163*** (0.011)
PISA 2012	0.030*** (0.005)	0.132*** (0.010)	0.170*** (0.010)	0.139*** (0.011)
PISA 2015	0.021*** (0.005)	0.112*** (0.010)	0.152*** (0.010)	0.119*** (0.011)
PISA 2018	-0.060*** (0.005)	0.037*** (0.011)	0.071*** (0.011)	0.023** (0.012)
Region (ref.: Rest of Spain)				
Castile and Leon	0.019*** (0.005)	0.109*** (0.009)	0.160*** (0.009)	0.118*** (0.009)
Catalonia	-0.111*** (0.003)	-0.054*** (0.010)	-0.065*** (0.010)	-0.159*** (0.011)
Basque Country	-0.115*** (0.005)	-0.190*** (0.013)	-0.078*** (0.012)	-0.205*** (0.013)
Constant	0.858*** (0.009)	-0.215*** (0.059)	0.118** (0.058)	0.183*** (0.062)
Observations	146,047	146,047	146,047	146,047
R-squared	0.176	0.288	0.304	0.204
Stock and Yogo (2005) test of weak instruments	-	65.225***	65.225***	65.225***

Note: Standard errors in parentheses. All OECD (2020c) recommended practices have been applied (final student weights, balanced repeated replication weights, plausible values). Dependent variable: Students' standardised competence scores. Two-stage least squares (2SLS). The instrument is the child's month of birth. The null hypothesis of the Stock and Yogo (2005) test of weak instruments is that the instrument is weak. Significance: \* significant at 0.1, \*\* significant at 0.5, \*\*\* significant at 0.01.

Source: Authors' own calculations.

TABLE 2 The influence of grade retention on students' standardised scores by gender, instrumental variables.

Variables	Male			Female				
	First stage	Reduced form		First stage	Reduced form			
		Reading	Mathematics		Science	Reading	Mathematics	Science
Repeater (ref.: non-repeater)	-	-1.573*** (0.092)	-1.793*** (0.091)	-	-1.483*** (0.103)	-1.508*** (0.101)	-	-1.645*** (0.109)
Month of birth (ref.: December)								
January	-0.106*** (0.008)	-	-	-	-0.084*** (0.007)	-	-	-
February	-0.100*** (0.008)	-	-	-	-0.080*** (0.008)	-	-	-
March	-0.085*** (0.008)	-	-	-	-0.085*** (0.007)	-	-	-
April	-0.090*** (0.008)	-	-	-	-0.070*** (0.007)	-	-	-
May	-0.083*** (0.008)	-	-	-	-0.060*** (0.007)	-	-	-
June	-0.067*** (0.008)	-	-	-	-0.041*** (0.007)	-	-	-
July	-0.044*** (0.008)	-	-	-	-0.056*** (0.007)	-	-	-
August	-0.041*** (0.008)	-	-	-	-0.024*** (0.007)	-	-	-
September	-0.042*** (0.008)	-	-	-	-0.031*** (0.007)	-	-	-
October	-0.034*** (0.008)	-	-	-	-0.037*** (0.007)	-	-	-

(Continues)

TABLE 2 (Continued)

Variables	Male			Female			
	Reduced form			Reduced form			
	First stage	Reading	Mathematics	Science	Reading	Mathematics	Science
November	-0.005 (0.008)	-	-	-	-0.016** (0.007)	-	-
Missing flag. Month of birth	-	-	-	-	-	-	-
Immigrant status (ref.: native)							
First-generation immigrant	0.157*** (0.007)	0.033* (0.019)	0.021 (0.019)	0.010 (0.020)	0.137*** (0.006)	-0.008 (0.018)	-0.067*** (0.018)
Second-generation immigrant	0.085*** (0.012)	0.101*** (0.025)	0.046* (0.025)	0.115*** (0.026)	0.042*** (0.012)	0.004 (0.022)	-0.039* (0.021)
Missing flag. Immigrant status	0.140*** (0.012)	-0.308*** (0.027)	-0.332*** (0.027)	-0.307*** (0.028)	0.138*** (0.014)	-0.280*** (0.029)	-0.321*** (0.028)
Father level of education (ref.: none)							
ISCED 1	-0.068*** (0.009)	-0.005 (0.018)	0.002 (0.018)	-0.020 (0.019)	-0.089*** (0.008)	0.038** (0.017)	0.025 (0.017)
ISCED 2	-0.059*** (0.009)	0.048*** (0.018)	0.061*** (0.017)	0.036* (0.018)	-0.118*** (0.008)	0.052*** (0.019)	0.049*** (0.020)
ISCED 3b, c	-0.090*** (0.012)	0.054** (0.024)	0.072*** (0.024)	0.027 (0.025)	-0.153*** (0.011)	0.016 (0.025)	0.037 (0.025)
ISCED 3a, 4	-0.093*** (0.009)	0.053*** (0.019)	0.044** (0.019)	0.036 (0.020)	-0.135*** (0.008)	0.068*** (0.021)	0.047** (0.020)
ISCED 5b	-0.082*** (0.009)	0.037* (0.019)	0.031 (0.019)	0.037* (0.020)	-0.144*** (0.009)	0.043** (0.022)	0.019 (0.021)
ISCED 5a, 6	-0.132*** (0.009)	0.065*** (0.021)	0.084*** (0.021)	0.051** (0.022)	-0.164*** (0.008)	0.064*** (0.023)	0.105*** (0.022)

TABLE 2 (Continued)

Variables	Male			Female				
	Reduced form			Reduced form				
	First stage	Reading	Mathematics	Science	First stage	Reading	Mathematics	Science
Missing flag. Father level of education	-0.027** (0.012)	-0.014 (0.023)	0.038* (0.022)	0.014 (0.024)	0.006 (0.010)	0.000 (0.019)	0.002 (0.019)	-0.014 (0.020)
Mother level of education (ref.: none)								
ISCED 1	-0.056*** (0.010)	0.199*** (0.019)	0.130*** (0.019)	0.167*** (0.020)	-0.057*** (0.009)	0.139*** (0.017)	0.049*** (0.017)	0.125*** (0.018)
ISCED 2	-0.064*** (0.010)	0.214*** (0.019)	0.172*** (0.019)	0.189*** (0.020)	-0.086*** (0.009)	0.137*** (0.018)	0.080*** (0.018)	0.143*** (0.019)
ISCED 3b, c	-0.107*** (0.013)	0.183*** (0.026)	0.132*** (0.026)	0.181*** (0.027)	-0.107*** (0.011)	0.184*** (0.023)	0.089*** (0.023)	0.180*** (0.025)
ISCED 3a, 4	-0.117*** (0.010)	0.179*** (0.022)	0.123*** (0.021)	0.159*** (0.023)	-0.120*** (0.009)	0.158*** (0.020)	0.081*** (0.020)	0.158*** (0.022)
ISCED 5b	-0.102*** (0.011)	0.135*** (0.022)	0.070*** (0.022)	0.139*** (0.023)	-0.105*** (0.009)	0.109*** (0.021)	0.023 (0.020)	0.126*** (0.022)
ISCED 5a, 6	-0.166*** (0.010)	0.158*** (0.024)	0.114*** (0.024)	0.136*** (0.026)	-0.152*** (0.009)	0.176*** (0.023)	0.134*** (0.022)	0.186*** (0.024)
Missing flag. Mother level of education	0.031** (0.014)	-0.067** (0.027)	-0.062** (0.027)	-0.038 (0.028)	-0.031** (0.014)	-0.070*** (0.026)	-0.061** (0.025)	-0.024 (0.027)
Number of books at home (ref.: 0–10 books)								
11–25 books	-0.083*** (0.007)	0.184*** (0.015)	0.142*** (0.015)	0.125*** (0.016)	-0.127*** (0.007)	0.114*** (0.018)	0.098*** (0.018)	0.049** (0.019)
26–100 books	-0.206*** (0.006)	0.316*** (0.022)	0.276*** (0.022)	0.247*** (0.024)	-0.252*** (0.007)	0.300*** (0.029)	0.296*** (0.028)	0.222*** (0.030)

(Continues)

TABLE 2 (Continued)

Variables	Male			Female				
	Reduced form			Reduced form				
	First stage	Reading	Mathematics	Science	First stage	Reading	Mathematics	Science
101-200 books	-0.279*** (0.007)	0.403*** (0.029)	0.373*** (0.029)	0.346*** (0.030)	-0.312*** (0.007)	0.426*** (0.035)	0.446*** (0.034)	0.365*** (0.037)
201-500 books	-0.311*** (0.007)	0.503*** (0.032)	0.487*** (0.032)	0.449*** (0.034)	-0.356*** (0.007)	0.502*** (0.039)	0.556*** (0.038)	0.464*** (0.041)
More than 500 books	-0.303*** (0.008)	0.448*** (0.032)	0.466*** (0.032)	0.443*** (0.034)	-0.346*** (0.008)	0.545*** (0.039)	0.609*** (0.038)	0.524*** (0.041)
Missing flag. Number of books at home	-0.100*** (0.017)	0.094*** (0.033)	0.093*** (0.033)	0.063* (0.035)	-0.166*** (0.021)	0.048 (0.042)	0.104** (0.041)	0.109** (0.044)
School funding (ref.: public)								
Private	-0.134*** (0.007)	0.056*** (0.018)	-0.001 (0.018)	-0.005 (0.019)	-0.083*** (0.006)	0.083*** (0.014)	0.048*** (0.014)	0.026* (0.015)
Semi-private	-0.098*** (0.004)	0.032*** (0.012)	-0.039*** (0.012)	-0.044*** (0.012)	-0.073*** (0.004)	0.054*** (0.010)	0.015 (0.010)	-0.022** (0.011)
Missing flag. School funding	-0.065*** (0.008)	-0.034** (0.017)	-0.070*** (0.017)	-0.100*** (0.018)	-0.048*** (0.008)	0.036** (0.015)	0.018 (0.015)	-0.038** (0.016)
PISA cycle (ref: PISA 2003)								
PISA 2006	0.096*** (0.008)	0.256*** (0.017)	0.285*** (0.017)	0.268*** (0.018)	0.091*** (0.007)	0.192*** (0.016)	0.261*** (0.016)	0.225*** (0.017)
PISA 2009	0.044*** (0.008)	0.192*** (0.015)	0.226*** (0.015)	0.174*** (0.016)	0.064*** (0.007)	0.111*** (0.015)	0.153*** (0.014)	0.148*** (0.015)
PISA 2012	0.008 (0.008)	0.171*** (0.015)	0.176*** (0.015)	0.136*** (0.016)	0.051 (0.007)	0.094*** (0.014)	0.154*** (0.014)	0.135*** (0.015)

TABLE 2 (Continued)

Variables	Male			Female				
	Reduced form			Reduced form				
	First stage	Reading	Mathematics	Science	First stage	Reading	Mathematics	Science
PISA 2015	0.005 (0.008)	0.214*** (0.015)	0.151*** (0.015)	0.132*** (0.016)	0.036*** (0.007)	0.009 (0.014)	0.147*** (0.013)	0.101*** (0.014)
PISA 2018	-0.086*** (0.008)	0.091*** (0.017)	0.018 (0.017)	-0.010 (0.018)	-0.036*** (0.007)	-0.018 (0.014)	0.119*** (0.013)	0.053*** (0.015)
Region (ref.: Rest of Spain)								
Castile and Leon	0.023*** (0.007)	0.141*** (0.013)	0.161*** (0.013)	0.131*** (0.014)	0.015** (0.006)	0.081*** (0.011)	0.162*** (0.011)	0.107*** (0.012)
Catalonia	-0.104*** (0.005)	-0.044*** (0.013)	-0.054*** (0.013)	-0.142*** (0.014)	-0.116*** (0.005)	-0.064*** (0.015)	-0.068*** (0.014)	-0.170*** (0.015)
Basque Country	-0.123*** (0.008)	-0.217*** (0.019)	-0.106*** (0.018)	-0.237*** (0.020)	-0.107*** (0.007)	-0.169*** (0.017)	-0.047*** (0.016)	-0.171*** (0.018)
Constant	0.839*** (0.014)	-0.252*** (0.076)	0.148** (0.075)	0.179** (0.080)	0.798*** (0.013)	-0.009 (0.081)	-0.262*** (0.078)	-0.084 (0.085)
Observations	72,985	72,985	72,985	72,985	73,061	73,061	73,061	73,061
R-squared	0.170	0.252	0.266	0.185	0.170	0.285	0.349	0.239
Stock and Yogo (2005) test of weak instruments	-	39.271***	39.271***	39.271***	-	28.733***	28.733***	28.733***

Note: Standard errors in parentheses. All OECD (2020c) recommended practices have been applied (final student weights, balanced repeated replication weights, plausible values). Dependent variable: Students' standardised competence scores. Two-Stage Least Squares (2SLS). The instrument is the child's month of birth. The null hypothesis of the Stock and Yogo (2005) test of weak instruments is that the instrument is weak. Significance: \* significant at 0.1, \*\* significant at 0.05, \*\*\* significant at 0.01.

Source: Authors' own calculations.

than girls'. While the competences of those male students who were held back a year decreases between 1.6 and 1.8 SD, girls' performance worsens between 1.5 and 1.7 SD.

In short, grade retention has a negative influence on Spanish students' competences, with boys being the worst off.

### 3 | DISCUSSION AND CONCLUSIONS

Throughout this paper, we have tried to analyse the effect that grade retention has on 15-year-old Spanish students' competences, using data from six PISA cycles (2003–2018). In order to go beyond a correlational analysis and obtain causal estimates, we have used a quasi-experimental methodology (instrumental variables). As mentioned, our results show that grade retention is negatively associated with students' competences between 1.5 and 1.7 SD, influencing boys' competences to a greater extent. For instance, holding back a year is negatively associated with boys' mathematics competence by 1.8 SD, while it presents a negative influence on girls' by 1.5 SD. These results may be due to the fact that classes in Spain are largely oriented towards content-based knowledge rather than skills (Jerrim et al., 2019; Marcenaro-Gutierrez & Vignoles, 2014) and, as a result, students who repeat grades will not only fail to master required knowledge but may also be at a greater disadvantage when it comes to taking PISA tests, which are competence-oriented.

In this sense, while grade retention has been widely used as a way of dealing with students who fail to keep up with their peers, it seems to perpetuate the failures of the Spanish education system. Far from offering students who are lagging behind new opportunities and support to achieve grade-level knowledge, grade retention seems to reinforce their shortcomings even more, as students fail to acquire the competences that they should have learned in the first-time round. And this inefficiency is compounded by the fact that students who repeat are not only more likely to drop out (Andrew, 2014; Glick & Sahn, 2010; Ou & Reynolds, 2010), but also to have lower self-esteem and academic self-concept (Lamote et al., 2014; Martin, 2011; Peixoto et al., 2016).

This may be due to the fact that students are often forced to retake the course under the same conditions without being offered any extra support or additional material (González-Betancor & López-Puig, 2016), and also that the emotional havoc that grade retention can cause may lead to problematic behaviours that hinder learning (Jimerson, 2001; Jimerson et al., 2005). The latter is particularly relevant, as boys tend to be more vulnerable to disruptive behaviour than girls (Pagani et al., 2001), which may explain why grade retention has a more detrimental influence on boys (Morrison & No, 2007).

In this context, early childhood interventions can be a very useful alternative to grade retention (García-Pérez et al., 2014). Early identification of those students who are at-risk can enable teachers to be aware of their situation and to use different communication strategies that can help them to improve their competences (Macfadyen & Dawson, 2010). To this aim, it could be beneficial to create intervention teams to analyse data about the individual students and, according to that information, to develop an 'individualised education plan', similar to those used with special needs' students. Alongside early intervention, automatic promotion is another alternative to be considered. However, and while this practice would allow to overcome the emotional havoc of detaching students from their peers, it should be applied carefully, as it may reduce students' interest in learning, their attendance and even teachers' effort (Ahmed & Mihiretie, 2015). In this sense, it seems that rather than going straight to the extremes, the most plausible measure is at some point between grade retention and automatic promotion, where students are given tailoring instruction to meet their needs (Reis et al., 2011; Wanner & Palmer, 2015).

An additional approach could be to provide students with supplemental instructional time after school, on the weekend or in summer programs. These reinforcement classes could be a much more valuable alternative to the private tuition used by a high proportion of families in Spain to improve the low performance of their children. Moreover, in order to provide high quality instruction to the low performing students, teachers may need specific professional instruction designed to acquire teaching skills and methodologies that could

help them diversify the way they approach these students. Additionally, permanent communication between teachers and parents is essential to identify the potential problems that may arise for these students with instructional needs.

Based on the relationship between grade retention and month of birth, another alternative could be following the policy that other countries have employed to make more flexible the enrolment of students in compulsory education (as described in 'Exhibit 2: Grade Assessed and Average Age of the Students Assessed in PIRLS 2016'; Mullis et al., 2017), which indicates that parents have discretion to make earlier or to delay their children's enrolment if they feel their children will benefit from being more mature.

In any case, our results should be interpreted with caution. There may still be unobservable variables that we have not controlled for and, moreover, by analysing only Spanish students, it is difficult to extrapolate our results to other educational systems. In this sense and, in order to overcome the low external validity, future research could focus on the international level.

## ACKNOWLEDGEMENTS

This work has been partly supported by FEDER funding (under Research Project PY20-00228-R); *Ministerio de Ciencia e Innovación* (under Research Project PID2020-119471RB-I00) and by the Andalusian Regional Government (SEJ-645), by the *Universidad de Málaga* (under Research Project B1-2022\_23), by the *Fundación Ramón Areces* and by the *Universidad de Málaga/CBUA* for the funding of Open Access Charge. We also acknowledge the scholarship FPU20/01509 of the *Ministerio de Universidades* and the training received from the *Programa de Doctorado en Economía y Empresa* of the *Universidad de Málaga*.

## CONFLICT OF INTEREST STATEMENT

The authors do not have any conflict of interest.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in PISA Data at <https://www.oecd.org/pisa/data/>.

## INFORMED CONSENT

Participants consented their participation.

## ORCID

Luis Alejandro Lopez-Agudo  <https://orcid.org/0000-0002-0906-3206>

María Ladrón de Guevara Rodríguez  <https://orcid.org/0000-0002-5087-422X>

Oscar David Marcenaro-Gutierrez  <https://orcid.org/0000-0003-0939-5064>

## ENDNOTES

<sup>1</sup>PISA uses 'literacy' as a synonym of competences.

<sup>2</sup>PISA defines reading literacy competence as 'understanding, using, reflecting on and engaging with written texts, in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society.' (OECD, 2019, p. 27), mathematics competence as 'capacity to formulate, employ and interpret mathematics in a variety of contexts. It includes reasoning mathematically and using mathematical concepts, procedures, facts and tools to describe, explain and predict phenomena. It assists individuals to recognise the role that mathematics plays in the world and to make the well-founded judgements and decisions needed by constructive, engaged and reflective citizens' (OECD, 2019, p. 75) and science competence as 'the ability to engage with science-related issues, and with the ideas of science, as a reflective citizen. A scientifically literate person is willing to engage in reasoned discourse about science and technology, which requires the competencies to explain phenomena scientifically, evaluate and design scientific enquiry, and interpret data and evidence scientifically' (OECD, 2019, p. 15).

- <sup>3</sup>PISA also assesses other competences such as financial literacy, problem-solving skills or the global competence. It also includes other background questionnaires such as parental, teacher, ICT, well-being, educational career questionnaires. However, the administration of these questionnaires has been performed irregularly by PISA cycles and countries, so the present study is focused on the competences and student information which remain fixed along PISA cycles.
- <sup>4</sup>BOE (2002, art. 17.3), BOE (2006, art. 20.2) and BOE (2013, art. 20.2).
- <sup>5</sup>BOE (2002, art. 29.3), BOE (2006, art. 28.5) and BOE (2013, art. 28.5).
- <sup>6</sup>It is important to highlight that, to the extent that PISA analyses students from the same age (15 years old) then repeater students are not lost from the sample due to grade retention (which could happen when sampling based on grade and not age). Therefore, potential positive bias in scores, which may happen when sampling based on students' grade, is avoided.
- <sup>7</sup>The Sargan test of overidentifying restrictions is relevant to check the validity of the instrument, whereas the Wu-Hausman test could provide information about the suitability of the 2SLS procedure. However, there is evidence indicating that, when weights are used in complex survey design data, these tests become invalid, so their conclusions could be misleading. This is indicated in Stata's help (Stata, 2024) for the command of 'estat overid' (Sargan test) and 'estat endogenous' (Wu-Hausman test), in the subcommand to force the use of survey weights (i.e. 'forceweights'): 'By default, these tests are conducted only after unweighted or frequency-weighted estimation. The reported critical values may be inappropriate for weighted data, so the user must determine whether the critical values are appropriate for a given application' (pp. 4–5) and 'By default, estat overid issues an error message if the previous estimation was conducted using awweights, pweights, or iweights. You can use the forceweights option to override this behaviour, though the test statistics may no longer have the expected  $\chi^2$  distributions' (p. 12). More discussion on this issue with complex survey design weighting and  $\chi^2$  test distributions can be found in Rao and Scott (1984). Therefore, to the extent that the results of these tests are difficult to believe, they have not been reported.
- <sup>8</sup>Tables A4 and A5 (Appendix A) have been replicated without parental education and results do not change; also, parental level of education and number of books at home have been, alternatively, replaced with the economic, social and cultural index created by the OECD, and results do not change. These estimations are available upon request to the authors.
- <sup>9</sup>The null hypothesis of the Stock and Yogo (2005) test of weak instruments is that the instrument is weak.

## REFERENCES

- Ahmed, A. Y., & Mihiretie, D. M. (2015). Primary school teachers and parents' views on automatic promotion practices and its implications for education quality. *International Journal of Educational Development*, 43, 90–99. <https://doi.org/10.1016/j.ijedudev.2015.05.003>
- Allen, C. S., Chen, Q., Willson, V. L., & Hughes, J. N. (2009). Quality of research design moderates effects of grade retention on achievement: A meta-analytic, multilevel analysis. *Educational Evaluation and Policy Analysis*, 31(4), 480–499. <https://doi.org/10.3102/0162373709352239>
- Andrew, M. (2014). The scarring effects of primary-grade retention? A study of cumulative advantage in the educational career. *Social Forces*, 93(2), 653–685. <https://doi.org/10.1093/sf/sou074>
- Barua, R., & Lang, K. (2016). School entry, educational attainment and quarter of birth: A cautionary tale of a local average treatment effect. *Journal of Human Capital*, 10(3), 347–376. <https://doi.org/10.1086/687599>
- Bedard, K., & Dhuey, E. (2006). The persistence of early childhood maturity: International evidence of long-run age effects. *The Quarterly Journal of Economics*, 121(4), 1437–1472. <https://doi.org/10.1093/qje/121.4.1437>
- Bernardi, F. (2014). Compensatory advantage as a mechanism of educational inequality. *Sociology of Education*, 87(2), 74–88. <https://doi.org/10.1177/0038040714524258>
- BOE. (2002). *Organic law 10/2002, 23rd December, of the quality of education (LOCE)*. Spain: N° 307, 24th December 2002, 45188–45220.
- BOE. (2006). *Organic law 2/2006, 3rd may, of education (LOE)*. Spain: N° 106, 4th May 2006, 17158–17207.
- BOE. (2013). *Organic law 8/2013, 9th December, for the improvement of the education quality (LOMCE)*. N° 295, 10th December 2013, 97858–97921. Spain.
- BOE. (2021). *Royal Decree 984/2021, 16th November, for the evaluation and promotion in primary education, as well as the evaluation, promotion and qualification in compulsory secondary education, baccalaureate and vocational training*. N° 275, 17th November 2021, 141583–141595. Spain.
- Calero, J., Choi, A., & Waisgrais, S. (2010). Determinants of the school failure risk in Spain: A multilevel logistic model approach to PISA-2006. *Revista de Educación, Extraordinary, Special Issue 2010*, 225–256.
- Calsamiglia, C., & Loviglio, A. (2019). Maturity and school outcomes in an inflexible system: Evidence from Catalonia. *SERIEs*, 11(4), 1–49. <https://doi.org/10.1007/s13209-019-0196-6>

- Carabaña, J. (2013). Repetición de curso y puntuaciones pisa ¿cuál causa cuál? In INEE (Eds.), In *PISA 2012: Programa para la evaluación internacional de los alumnos. Informe español. Volumen II: Análisis secundario* (pp. 32–66). Ministerio de Educación, Cultura y Deporte, Instituto Nacional de Evaluación Educativa.
- Chen, X., Liu, C., Zhang, L., Shi, Y., & 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(6), 544–559. <https://doi.org/10.1016/j.ijedudev.2009.12.002>
- Cook, P. J., & Kang, S. (2016). Birthdays, schooling, and crime: Regression-discontinuity analysis of school performance, delinquency, dropout, and crime initiation. *American Economic Journal: Applied Economics*, 8(1), 33–57. <https://doi.org/10.1257/app.20140323>
- Cordero, J. M., & Pedraja, F. (2019). The effect of financial education training on the financial literacy of Spanish students in PISA. *Applied Economics*, 51(16), 1679–1693. <https://doi.org/10.1080/00036846.2018.1528336>
- Dhuey, E., Figlio, D., Karbownik, K., & Roth, J. (2019). School starting age and cognitive development. *Journal of Policy Analysis and Management*, 38(9), 538–578. <https://doi.org/10.1002/pam.22135>
- Diris, R. (2017). Don't hold back? The effect of grade retention on student achievement. *Education Finance and Policy*, 12(3), 312–341. [https://doi.org/10.1162/EDFP\\_a\\_00203](https://doi.org/10.1162/EDFP_a_00203)
- Eide, E. R., & Showalter, M. H. (2001). The effect of grade retention on educational and labor market outcomes. *Economics of Education Review*, 20(6), 563–576. [https://doi.org/10.1016/S0272-7757\(00\)00041-8](https://doi.org/10.1016/S0272-7757(00)00041-8)
- Fiorini, M., & Stevens, K. (2014). *Assessing the Monotonicity assumption in IV and fuzzy RD designs*. The University of Sydney Economics. Working Paper Series, 13, 1–52.
- García-Pérez, J. I., Hidalgo-Hidalgo, M., & Robles-Zurita, J. A. (2014). Does grade retention affect achievement? Some evidence from Spain. *Applied Economics*, 46(12), 1373–1392. <https://doi.org/10.1080/00036846.2013.872761>
- Glick, P., & Sahn, D. E. (2010). Early academic performance, grade repetition, and school attainment in Senegal: A panel data analysis. *The World Bank Economic Review*, 24(1), 93–120. <https://doi.org/10.2307/40647422>
- González-Betancor, S. M., & López-Puig, A. J. (2016). Grade retention in primary education is associated with quarter of birth and socioeconomic status. *PLoS One*, 11(11), e0166431. <https://doi.org/10.1371/journal.pone.0166431>
- Goos, M., Damme, J., Onghena, P., Petry, K., & Bilde, J. (2013). First-grade retention in the Flemish educational context: Effects on children's academic growth, psychosocial growth, and school career throughout primary education. *Journal of School Psychology*, 51(3), 323–347. <https://doi.org/10.1016/j.jsp.2013.03.002>
- Goos, M., Pipa, J., & Peixoto, F. (2021). Effectiveness of grade retention: A systematic review and meta-analysis. *Educational Research Review*, 34(1), 100401. <https://doi.org/10.1016/j.edurev.2021.100401>
- Gortazar, L. (2019). *¿Favorece el sistema educativo español la igualdad de oportunidades?* FEDEA, Estudios sobre la Economía Española no. 2019/17, Madrid.
- Hanchane, S., & Mostafa, T. (2010). *Endogeneity problems in multilevel estimation of education production functions: An analysis using PISA data*. LLAKES research paper 14, 1–45. Centre for Learning and Life Chances in Knowledge Economies and Societies.
- Hwang, S., & Cappella, E. (2018). Rethinking early elementary grade retention: Examining long-term academic and psychosocial outcomes. *Journal of Research on Educational Effectiveness*, 11(4), 559–587. <https://doi.org/10.1080/19345747.2018.1496500>
- Inglés, C., González, C., García-Fernández, J., Vicent Juan, M., & Martínez-Monteaigudo, M. (2015). Current status of research on school refusal. *European Journal of Education and Psychology*, 8(1), 37–52. <https://doi.org/10.1016/j.ejeps.2015.10.005>
- Jacob, B., & Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *The Review of Economics and Statistics*, 86(1), 226–244. <https://doi.org/10.1162/003465304323023778>
- Jerrim, J., Lopez-Agudo, L., Marcenaro-Gutierrez, O., & Shure, N. (2017). What happens when econometrics and psychometrics collide? An example using the PISA data. *Economics of Education Review*, 61, 51–58. <https://doi.org/10.1016/j.edurev.2017.09.007>
- Jerrim, J., Lopez-Agudo, L. A., & Marcenaro-Gutierrez, O. D. (2019). The relationship between homework and the academic progress of children in Spain during compulsory elementary education: A twin fixed-effects approach. *British Educational Research Journal*, 45(5), 1021–1049. <https://doi.org/10.1002/berj.3549>
- Jerrim, J., Lopez-Agudo, L. A., & Marcenaro-Gutierrez, O. D. (2022). Grade retention and school entry age in Spain: A structural problem. *Educational Assessment, Evaluation and Accountability*, 34, 331–359. <https://doi.org/10.1007/s11092-021-09375-7>
- Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review*, 30(3), 420–437. <https://doi.org/10.1080/02796015.2001.12086124>
- Jimerson, S. R., & Kaufman, A. M. (2003). Reading, writing, and retention: A primer on grade retention research. *The Reading Teacher*, 56(7), 622–635.

- Jimerson, S. R., Pletcher, S. M. W., Graydon, K., Schnurr, B. L., Nickerson, A. B., & Kundert, D. K. (2005). Beyond grade retention and social promotion: Promoting the social and academic competence of students. *Psychology in the Schools*, 43(1), 85–97. <https://doi.org/10.1002/pits.20132>
- Klapproth, F., Schaltz, P., Brunner, M., Keller, U., Fischbach, A., Ugen, S., & Martin, R. (2016). Short-term and medium-term effects of grade retention in secondary school on academic achievement and psychosocial outcome variables. *Learning and Individual Differences*, 50, 182–194. <https://doi.org/10.1016/j.lindif.2016.08.014>
- Lamote, C., Pinxten, M., Van den Noortgate, W., & Van Damme, J. (2014). Is the cure worse than the disease? A longitudinal study on the effect of grade retention in secondary education on achievement and academic self-concept. *Educational Studies*, 40, 496–514. <https://doi.org/10.1080/03055698.2014.936828>
- Lorence, J. (2014). Third-grade retention and reading achievement in Texas: A nine year panel study. *Social Science Research*, 48, 1–19. <https://doi.org/10.1016/j.ssresearch.2014.05.001>
- Lounkaew, K. (2013). Explaining urban–rural differences in educational achievement in Thailand: Evidence from PISA literacy data. *Economics of Education Review*, 37, 213–225. <https://doi.org/10.1016/j.econedurev.2013.09.003>
- Macfadyen, L., & Dawson, S. (2010). Mining LMS data to develop an “early warning system” for educators: A proof of concept. *Computers & Education*, 54(2), 588–599. <https://doi.org/10.1016/j.compedu.2009.09.008>
- Mahjoub, M. B. (2017). The treatment effect of grade repetitions. *Education Economics*, 25(4), 1–15. <https://doi.org/10.1080/09645292.2017.1283006>
- Manacorda, M. (2012). The cost of grade retention. *Review of Economics and Statistics*, 94(2), 596–606. [https://doi.org/10.1162/rest\\_a\\_00165](https://doi.org/10.1162/rest_a_00165)
- Marcenaro-Gutierrez, O. D., & Vignoles, A. (2014). A comparison of teacher and test-based assessment for Spanish primary and secondary students. *Educational Research*, 57(1), 1–21. <https://doi.org/10.1080/00131881.2014.983720>
- Martin, A. (2011). Holding back and holding behind: Grade retention and students' non-academic and academic outcomes. *British Educational Research Journal*, 37(5), 739–763. <https://doi.org/10.1080/01411926.2010.490874>
- Martin, A. J. (2009). Age appropriateness and motivation, engagement, and performance in high school: Effects of age within cohort, grade retention, and delayed school entry. *Journal of Educational Psychology*, 101(1), 101–114. <https://doi.org/10.1037/a0013100>
- Mattison, A., Raffaele, L. M., Dedrick, R., Dickinson, S., Wingate, E., & Hanks, C. (2018). Early elementary teacher ratings of behavior as predictors of grade retention: Race, gender, and socioeconomic status as potential moderators. *Psychology in the Schools*, 55(10), 1171–1187. <https://doi.org/10.1002/pits.22192>
- Morrison, K., & No, A. (2007). Does repeating a year improve performance? The case of teaching English. *Educational Studies*, 33(3), 353–371. <https://doi.org/10.1080/03055690701423333>
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). *PIRLS 2016 international results in Reading*. Boston College.
- OECD. (2014). *PISA 2012 technical report*. OECD Publishing.
- OECD. (2017). *PISA 2015 assessment and analytical framework: Science, Reading, mathematics, financial literacy and collaborative problem solving, revised edition*. PISA, OECD Publishing. <https://doi.org/10.1787/9789264281820-en>
- OECD. (2019). *PISA 2018 assessment and analytical framework*. PISA, OECD Publishing. <https://doi.org/10.1787/b25feab8-en>
- OECD. (2020a). *PISA 2018 results (volume V): Effective policies, successful schools*. OECD Publishing.
- OECD. (2020b). *PISA 2018 results. Are students ready to thrive in an interconnected world? Volume VI*. OECD Publishing. <https://doi.org/10.1787/d5f68679-en>
- OECD. (2020c). *PISA 2018 technical report*. OECD Publishing.
- Ou, S.-R., & Reynolds, A. J. (2010). Grade retention, postsecondary education, and public aid receipt. *Educational Evaluation and Policy Analysis*, 32(1), 118–139. <https://doi.org/10.3102/0162373709354334>
- Pagani, L., Tremblay, R., Vitaro, F., Boulerice, B., & McDuff, P. (2001). Effects of grade retention on academic performance and behavioral development. *Development and Psychopathology*, 13(2), 297–315. <https://doi.org/10.1017/S0954579401002061>
- Pedraja-Chaparro, F., Santín, D., & Simancas, R. (2015). Determinants of grade retention in France and Spain: Does birth month matter? *Journal of Policy Modeling*, 37(5), 820–834. <https://doi.org/10.1016/j.jpolmod.2015.04.004>
- Peixoto, F., Monteiro, V., Mata, L., Sanches, C., Pipa, J., & Almeida, L. (2016). “To be or not to be retained ... That's the question!” retention, self-esteem, self-concept, achievement goals, and grades. *Frontiers in Psychology*, 7(742), 1550. <https://doi.org/10.3389/fpsyg.2016.01550>
- Ponzo, M., & Scoppa, V. (2014). The long-lasting effects of school entry age: Evidence from Italian students. *Journal of Policy Modeling*, 36(3), 578–599. <https://doi.org/10.1016/j.jpolmod.2014.04.001>
- Rao, J. N. K., & Scott, A. J. (1984). On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data. *The Annals of Statistics*, 12(1), 46–60. <https://doi.org/10.2307/2241033>

- Reis, S., Mccoach, D. B., Little, C., Muller, L., & Kaniskan, R. (2011). The effects of differentiated instruction and enrichment pedagogy on Reading achievement in five elementary schools. *American Educational Research Journal*, 48(2), 462–501. <https://doi.org/10.3102/0002831210382891>
- Robles-Piña, R. (2011). Depression and self-concept: Personality traits or coping styles in reaction to school retention of Hispanic adolescents. *Depression Research and Treatment*, 2011(1), 151469. <https://doi.org/10.1155/2011/151469>
- Schwerdt, G., West, M., & Winters, M. (2017). The effects of test-based retention on student outcomes over time: Regression discontinuity evidence from Florida. *Journal of Public Economics*, 152(4), 154–169. <https://doi.org/10.1016/j.jpubeco.2017.06.004>
- Sprietsma, M. (2010). The effect of relative age in the first grade of primary school on long-term scholastic results: International comparative evidence using PISA 2003. *Education Economics*, 18(1), 1–32. <https://doi.org/10.2139/ssrn.997221>
- Stata. (2024). *ivregress postestimation – Postestimation tools for ivregress*. [https://www.stata.com/manuals13/rivregress\\_postestimation.pdf](https://www.stata.com/manuals13/rivregress_postestimation.pdf)
- Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. In D. W. K. Andrews & J. H. Stock (Eds.), *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg* (pp. 80–108). Cambridge University Press.
- Taniguchi, K. (2015). Determinants of grade repetition in primary school in sub-Saharan Africa: An event history analysis for rural Malawi. *International Journal of Educational Development*, 45, 98–111. <https://doi.org/10.1016/j.ijedudev.2015.09.014>
- Tingle, L. R., Schoeneberger, J., & Algozzine, B. (2012). Does grade retention make a difference? *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 85(5), 179–185. <https://doi.org/10.1080/00098655.2012.679325>
- Wanner, T., & Palmer, E. (2015). Personalising learning: Exploring student and teacher perceptions about flexible learning and assessment in a flipped university course. *Computers & Education*, 88, 354–369. <https://doi.org/10.1016/j.compedu.2015.07.008>
- Wu, W., West, S., & Hughes, J. (2008). Short-term effects of grade retention on the growth rate of woodcock-Johnson III broad math and Reading scores. *Journal of School Psychology*, 46(1), 85–105. <https://doi.org/10.1016/j.jsp.2007.01.003>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Lopez-Agudo, L. A., de Guevara Rodríguez, M. L., & Marcenaro-Gutierrez, O. D. (2024). The influence of grade retention on students' competences in Spain. *European Journal of Education*, 00, e12736. <https://doi.org/10.1111/ejed.12736>