JOURNAL OF ECONOMIC SURVEYS

doi: 10.1111/joes.12406

EFFECTS OF GRADE RETENTION POLICIES: A LITERATURE REVIEW OF EMPIRICAL STUDIES APPLYING CAUSAL INFERENCE

Javier Valbuena*

University of Zaragoza

Mauro Mediavilla

University of Valencia

Álvaro Choi

University of Barcelona

María Gil

University Autónoma of Madrid

Abstract. The identification of the causal effects of grade retention policies is of enormous relevance for researchers and policymakers alike. Taking advantage of the availability of more detailed longitudinal datasets, researchers have been able to apply different identification strategies that address the classical problems of selection bias and unobserved heterogeneity that have plagued previous studies on the effect of retention. We present a systematic literature review of empirical studies aiming to unveil the causal effects of retention. This study underlines the need to consider and evaluate different kinds of grade retention polices as their effects vary depending on several dimensions (such as timing of the policy, comparison groups, length of the effects or institutional settings). According to the results of our review, we conclude that grade retention is unlikely to be an efficient policy as the costs associated to the policy can easily outweigh the potential (weak) benefits of retention. It is therefore necessary to consider alternative policies to retention, or policies that can be used in combination with it, in order to enhance the performance of low achievers, in particular those students at risk characterized by a low ability profile.

Keywords. Academic achievement; Causal inference; Grade retention effects; Literature review

1. Introduction

Grade retention policies are applied in many OECD countries as the primary method for enhancing the academic performance of low achievers, albeit with considerable variation in their grade retention rates. However, most recent available evidence seems to stress its negative effect on academic performance and

*Corresponding author contact email: jvalbuena@unizar.es; Tel: +34 976761837 (Ext. 841837).

labour market outcomes (Jacob *et al*., 2009; Glick and Sahn, 2010; Brodaty *et al*., 2013; Andrew, 2014; Belot *et al*., 2014; Fruehwirth *et al*., 2016; Gary-Bobo *et al*., 2016; Diris, 2017; Cockx *et al*., 2019).

The main reason for implementing the practice of grade retention is to act as a disincentive for the students' poor performance, that is, their inability to pass a certain number of subjects, as established by the syllabus. Several arguments have been used in defence of its use. First, it provides students with time to mature; second, it serves the purpose of establishing minimum academic requirements in order to advance to the next grade; finally, it aims to enhance overall performance by transmitting to students a culture of effort. In this sense, the claim is that it works as a deterrent to low performance (Manacorda, 2012). Overall, it is assumed that retention can improve the academic performance of low achievers by exposing them to an additional year of teaching so they catch up in terms of curriculum requirements. Nevertheless, grade retention remains a controversial measure, as the international empirical evidence seems to go against these arguments. Only in very specific institutional settings, and combined with alternative remedial measures (e.g. summer school, instructional support and better quality teachers) do results tend to be positive in the short run (i.e. high stake testing policies in Chicago and Florida).

Those opposed to grade retention emphasize its inefficacy (Jimerson *et al*., 2002), and its high cost (one of the most expensive educational policies, in fact): £6,000 per pupil per year in England (Education Endowment Foundation, based on the compilation of different meta-analysis), higher than $12 billion per year in the USA (West, 2012); 10% to 12% of total expenditure on primary and secondary education in Belgium, Spain, the Netherlands and Portugal, 5% to 10% in Brazil, Germany and Italy (OECD, 2011). Besides, there is a negative impact on student motivation, given that those required to repeat the grade are separated from their friends, obliged to retake the subjects not only they failed but those they passed, and to suffer the potential stigma of being labelled 'slow' students (Martin, 2011). Moreover, the practice may also generate discipline issues in schools (Crothers *et al*., 2010) and inequality issues are enlarged when applying grade retention (Spruyt, 2008; Kloosterman and De Graaf, 2010). All in all, it might negatively affect academic performance and increase the probability of school dropout (Holmes, 1989; Jacob and Lefgren, 2009). Therefore, the efficacy of grade retention is not only of academic interest, but also of enormous policy relevance given its consequences for students and schools alike (De Witte *et al*., 2013).

Our study contributes by providing a complete literature review of empirical studies using causal inference techniques. This is crucial in terms of policy implications as the potential unobservable heterogeneity associated to the non-random selection of the students into grade retention acts as a confounding factor which, when ignored, would impose severe limitations on the evaluation of the retention policy. The survey includes 42 papers in the timeframe from 2001 to 2020 that credibly address the identification challenges by making use of a variety of methodologies. These include experimental techniques and a set of quasi-experimental approaches, the most relevant being instrumental variables (IVs), difference in differences (DiD), regression discontinuity designs (RDD) and structural models.

Furthermore, we present our results from different angles classifying the evidence according to the outcome of interest evaluated in each of the studies. We first focus on the effect of grade retention on student outcomes considering the educational level where the policy is applied (i.e. kindergarten, primary school and secondary school), and whether the policy shows lasting effects. Second, we evaluate the resulting evidence of the policy on other educational attainment (e.g. graduation rates and dropout) and labour market outcomes. Third, we also consider to what extent the results are affected by particular institutional settings analysing the effects when retention is combined with alternative remedial measures. Finally, we scaffold our study by providing evidence on the effect of alternative policies to grade retention based on early intervention and individual treatment.

The final aim of this study is to offer policymakers a solid body of proofs on the effects of this measure, and on the possible and already tested alternatives that could remedy the negative effects of grade retention. Indeed, most of the evidence compiled suggests the negative impact of grade retention

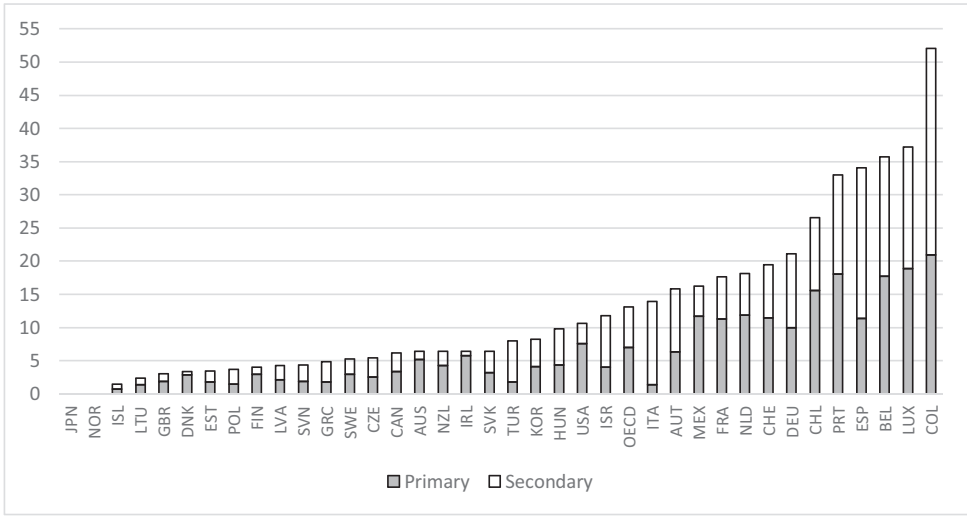

Figure 1. Students Who Have Repeated at Least One Grade at Age 15 (%). OECD Countries, Programme for International Student Assessment (PISA) 2018.
Source: Self-elaboration from OECD data.

on different aspects of educational outcomes, which calls for the necessity to consider alternative policies especially for those students with a lower academic performance.

The remainder of the paper is organized as follows. Section 2 presents the context and relevance of grade retention as an educational policy. Section 3 describes the characteristics of students affected by the policy. Section 4 presents the methodology of the systematic literature review conducted, distinguishing categories associated to different methodologies and robust econometric approaches. Section 5 presents the evidence on the causal effects of grade retention on a set of educational and labour market outcomes as well as on the combined causal effects of retention and alternative measures. Section 6 briefly comments on potential alternative policies to grade retention. Section 7 discusses the policy implications related to the contents of the empirical studies surveyed in the previous sections. Finally, Section 8 draws some conclusions.

2. Contextual Analysis of the Relevance of Grade Retention

Although the available empirical evidence tends to underline the negative effects of grade retention, several OECD countries still apply it as their main policy for tackling low academic performance (see Figure 1). The main reason to make a student repeat a grade is their deficient academic progress, but absenteeism or misbehaviour are other reasons to apply grade retention. While countries such as Japan, Iceland or Norway prefer automatic promotion, others, such as Belgium and Portugal, use grade retention intensively. Grade repetition is preferably applied at the primary school level in countries such as France, Mexico, Ireland or the Netherlands, while in countries such as Italy, Spain or Colombia, most students repeat at the secondary school level.

Focusing on the EU, in most of the member states, there is a very specific normative regarding grade retention implementation (Borodankova and Coutinho, 2011), applicable for students with learning difficulties or those who have not reached the learning objectives of the course. Most countries apply

grade retention, although with some limitations to its extent – such as restraining its application in primary school, or the possible number of courses that can be repeated, or even specifying certain courses that cannot be repeated. Other countries, such as the UK, do not have a specific regulation: they apply instead the principle that education must be adequate to the age, capacity and skills of the student. But even countries with very similar regulations end up applying grade retention in very different ways. For instance, Spain and Luxembourg have much higher repetition rates than Cyprus or Slovakia in the primary level. Moreover, some countries in which grade retention is allowed do not apply it intensively: in this sense, Greece follows a complex process to decide whether a student needs to repeat or not; Italy requires the unanimity of all teachers; in Cyprus, the final decision is up to the inspector; Norway and Iceland generally apply automatic promotion, while Bulgaria or Liechtenstein do so in the primary level. Regarding the opinion of parents, those countries in which they are taken into account seem to have lower rates of repetition (for instance, Denmark and Sweden) although the final decision is generally taken at the school level.

Summarizing, the fact is that factors beyond regulation, such as tradition, cultural factors and social beliefs regarding the benefits and effectiveness of grade retention seem to play an important role in accounting for the differences in its application. Belgium, Luxembourg, Spain, Portugal, the Netherlands and France are examples of an extensive application of grade retention for these reasons (Goos *et al*., 2013a). Thus, the application of grade retention varies substantially across countries and over time which seems to be in line with the idea of questioning the measure per se, rather than trying to propose marginal changes in regulation.

In the USA, there is no federal law which regulates grade retention, although the 2001 'No Child Left Behind Act' reopened the possibility to adapt the federal regulation in terms of the requirements for the accountability systems, which is used to design the grade retention policy of many states and districts (Center for Mental Health in Schools at UCLA, 2008 Update). Therefore, states apply different approaches to promote or retain students. Regulation for retention is authorized in some states (via different channels), while 17 states have no specific regulation for grade retention. Social promotion conditioned on demonstrating proficiency is regulated in 12 states. Finally, 18 states use assessment to determine whether the student promotes or stays in the same grade. Using standardized tests in order to determine individual promotion or not, while it may seem to be an objective levelling measure, has nevertheless been contested by many studies. In fact, schools can adopt strategies to retain certain students in a grade to avoid a lessening in test score averages (Jacob, 2005; Werner, 2013) or low-scorers may be separated from their classmates and directed to special education (Haney, 2000).

Comprehensive accountability policies are designed with the aim of raising academic achievement. These policies usually combine grade retention with remedial interventions aimed at preserving the incentive effect associated with repetition, while attenuating its negative consequences. On the other hand, accountability can also leave disadvantaged students behind, as this system does not provide incentives to pay more attention to this type of students (Neal and Schanzenbach, 2010). The empirical evidence available shows that the effects of accountability systems on academic performance are mixed. Roderick and Nagaoka (2005) found no substantial positive effects for third graders in Chicago Public Schools (CPS), and negative effects for sixth graders. Similarly, Jacob and Lefgren (2004), also for CPS, showed that grade retention increases achievement in third graders, but not so in sixth graders. Conversely, more stringent accountability standards, in the case of Florida, appear to be associated with higher test scores (Winters and Greene, 2012).

3. Characterization of Students Suffering Grade Retention

There is a substantial literature on the determinants associated with the decision to retain children. This literature has focused on academic and non-academic factors, students' initial school skills, as well as

socioeconomic and family background factors, and time invariant individual characteristics such as age and gender effects, or the immigrant status of the student. Broadly speaking, greater grade repetition rates are associated to lower academic performance, to less motivation and to misbehaviour and certain personal characteristics of students (OECD, 2016). In what follows, we revise the main studies and their corresponding results in detail.

Student academic performance has been identified by the literature as one of the main predictors of grade retention considering both developed and developing countries. Children showing poor academic achievement (i.e. reading and writing) at the beginning of their schooling have a higher probability of repeating a grade during subsequent years in the USA (Ferguson *et al*., 2001; Bali *et al*., 2005; Frey, 2005; Wilson and Hughes, 2009), China (Chen *et al*., 2010), South Africa (Liddell *et al*., 2001) and Brazil (Gomes Neto and Hanushek, 1994). Another branch of the literature looks at the importance of human capital accumulation prior to school entry, even kindergarten, showing that early acquired skills do prepare children for schooling and reduce the probability of retention (Ferguson *et al*., 2001; Wilson and Hughes, 2009; Cordero *et al*., 2014; Carabaña, 2015; Agasisti and Cordero, 2017). This association is particularly strong for children of high-income parents (Elder and Lubotsky, 2009).

There is another group of authors that have approached the determinants of repetition focusing on other relevant individual characteristics, particularly looking at the relative age of children within their cohort. All these studies share a common relevant finding: the age at which a child starts school is negatively associated with the tendency to repeat grades (Corman, 2003; Guevremont *et al*., 2007; Chen *et al*., 2010; Verachtert *et al*., 2010; Pedraja *et al*., 2015). There is also an effect of relative age on medium- and long-term test scores, as older children show better academic performance and are more likely to attend college (Bedard and Dhuey, 2006; Pereira and Reis, 2014). In particular, relative age effects are a consequence of grade retention spreading pupils across grades (Sprietsma, 2010). These results hold even for kindergarten entrance – 'academic redshirting' (Elder and Lubotsky, 2009). Gender effects have also been analysed by the literature. The general finding is that boys are more likely to be retained (Frey, 2005; Chen *et al*., 2010; Pereira and Reis, 2014). Finally, the studies looking at the effects of immigrant status upon grade retention show that the risk of foreign students being retained is higher compared with their native counterparts (Frey, 2005; Bonvin *et al*., 2008; Cordero *et al*., 2014). Furthermore, immigrants of first and second generations are at risk of retention for a greater portion of their school careers than are non-immigrants but are significantly less likely to experience grade retention than their third-generation peers (Tillman *et al*., 2006).

However, grade repetition does not depend only on the pupil's actual academic performance but also on non-academic factors. Studies that analyse determinants of retention beyond academic achievement highlight the impact of teachers' attitudes and evaluations (Bonvin *et al*., 2008; Wynn, 2010) and reveal that students showing maladaptive behaviour or characterized as less confident, less self-assured and less engaging, and have a significantly higher probability of being retained than their academically similar peers (Jimerson *et al*., 2007). The degree of responsibility of students also plays a role in repetition (Pérez *et al*., 2009). Retention rates also depend significantly on socioeconomic factors as measured by family and neighbourhood characteristics (Corman, 2003; Frey, 2005; Jimerson *et al*., 2007) and parental education (Gomes Neto and Hanushek, 1994; Ferguson *et al*., 2001; Chen *et al*., 2010). Higher probabilities of being retained are associated to children coming from more disadvantaged backgrounds and having lower educated parents. Besides, socially disadvantaged students tend to make less ambitious educational choices, thus increasing socioeconomic and educational inequality (Spruyt, 2008; Kloosterman and De Graaf, 2010). Finally, the implication of parents in their children's education has a positive effect in avoiding grade repetition (Pérez-Díaz *et al*., 2001; Pérez *et al*., 2009).

There are a few studies approaching the analysis of grade retention from a cross-country perspective. General results for OECD countries indicate that traditions and societal beliefs regarding the benefits of grade retention play an important role in explaining international differences (Eurydice, 2011; Goos *et al*., 2013a), as do attitudes towards school (Ikeda and García, 2014). Besides, socioeconomic factors, student

composition of the school and early acquired skills are relevant for accounting for grade retention patterns within the EU (Agasisti and Cordero, 2017). Moreover, institutional factors may play an important role in the observed differences between countries (Pereira and Reis, 2014).

4. Meta-Analysis of Studies

4.1 *Method and Search Strategy*

In this section, we will review empirical studies that do control for endogeneity by means of applying robust methodologies in order to estimate the causal effect of grade retention. With this purpose, we set the following set of inclusion criteria to select studies for this review: (1) The study is published between 2001 and 2020 in peer-reviewed academic journals in the English language. We will also include relevant working papers based on whether they apply a rigorous methodological approach. (2) We consider both empirical and theoretical studies with empirical applications, whereas purely theoretical, conceptual and case studies are excluded. (3) The study deals with the causal effect of grade retention (considering the application of the educational policy during kindergarten, primary or secondary school) on educational or employment outcomes. Thus, papers using traditional econometric methods, such as least squares, whose estimated effects are highly unlikely to reveal causal implications, are also excluded. Table 1 provides a descriptive analysis of the studies surveyed by means of their most relevant characteristics.

We carried out our literature search from March 2019 to September 2020. The search was conducted in the following electronic databases: Educational Resources Information Center (ERIC), ISI Web of Science (WoS), Scopus, Econlit, SocINDEX and Google Scholar. ERIC, as the largest educational database worldwide providing access to approximately 1000 scientific journals, was used as the main search engine. The rest of the databases were used in order to add any articles that we might have missed to our results. We performed a systematic computerized search strategy using a wide range of search keywords, namely, 'grade retention', 'repetition', 'repeating', 'hold back', combined with 'kindergarten', 'primary school', 'secondary school', 'high school', 'college' and 'effects', 'achievement', 'test scores', 'drop out' or 'educational outcome', 'employment outcome'. We applied filters considering the type of document (e.g. papers, literature reviews, working papers and thesis) and by publication date starting from 2001 up to the year 2020. Our initial search identified more than 150 studies. All registers were extracted using the reference manager Mendeley, eliminating duplicates and selecting those that fulfilled our set of selection criteria after reading the title and abstract. Finally, once we had reviewed their content reading the full text of the articles, the number of studies was significantly reduced by eliminating all studies that did not use a rigorous methodological approach to identify the causal effect of grade retention. The final selection included 42 studies.

Table A1 describes the main characteristics of the studies meeting our set of inclusion criteria. Each register includes information regarding the country/countries under analysis, the database employed, the methodological approach, an overview of the main results presented in the paper as well as the type of data use to develop the study (longitudinal or cross-sectional).

4.2 *Methodologies*

There is a substantial literature on the effects of grade retention. Studies looking at the impact of retention have traditionally been conducted in educational research, but most of these early contributions that showed a clear negative effect of retention did not address endogeneity or selection problems in a credible way (see meta-analysis by Holmes and Matthews, 1984; Holmes, 1989; Jimerson, 2001; Xia and Kirby, 2009). Therefore, the presence of unobservable heterogeneity is a confounding factor that makes it difficult to draw inferences from this body of research. Most likely the decision to retain (promote) one

Table 1. Studies that Do Control for Endogeneity: Descriptive Analysis[a] Period: 2001–2020 ($N = 42$).

Descriptive categories			Number of studies
Types and groups of comparison (one paper with the two approach)[b]	Same-age		9
	Same-grade		34
Main statistical approach	Hierarchical models		4
	IV		6
	DiD		6
	RDD		7
	Structural models		2
	Matching methods		4
	Growth curve analysis		3
	Others approaches		10
Grade retention: level where is done	Kindergarten		2
	Primary		27
	Secondary		9
	Different levels		4
Outcome (some articles have more than one outcome)	Educational achievement	Maths	18
		Reading	23
		Science	3
	Test scores		7
	Reach the secondary school		4
	Dropout effect		6
	Others educational outcomes		4
	Employment outcomes		3
Year of publication	2001–2008		6
	2009–2020		36

Source: Own elaboration.
[a]Some of the studies from 2001 to 2008 included in our revision have already been surveyed in the paper by Xia and Kirby (2009), that is, Eide and Showalter (2001); Jacob and Lefgren (2004) and Green and Winters (2007). We have decided to start our literature review from 2001 for completeness, as the study by Eide and Showalter (2001) is considered to be the first attempt to estimate the causal effect of grade retention (Brodaty et al, 2013; Gary-Bobo *et al*., 2016). Moreover, our review includes studies from 2001 to 2008 that were not covered in Xia and Kirby (2009). In any case, the majority of the studies surveyed in this paper date from 2009 onwards (36 out of 42).
[b]The paper by Schwerdt *et al*. (2017) uses both same-age and same-grade comparisons.

student is made based on characteristics unobserved by the researcher. Thus, these studies are likely to suffer from serious selection bias as grade repeaters are a selected population that may differ substantially from potential comparison groups, such as promoted students (e.g. retained students are more likely to have lower innate ability and weaker social background than promoted students). Consequently, this literature is informative about the correlative relationship of repetition and different measures of educational achievement, but it should not be the basis for a discussion on policy implications.

In their meta-analysis, Allen *et al*. (2009) highlighted the relevance of methodological approaches in reaching conclusions regarding the effects of grade retention on achievement. They concluded that the overall average effect of retention was strongly related to the design quality of the studies under

consideration. In particular, they found that rigorous methodological approaches moderated the (negative) effect sizes, suggesting that retention may exert a slightly negative or even neutral effect on student achievement, thus indicating that retention does not benefit the student.

In this section, we present a systematic literature review regarding the impact of grade retention on student outcomes. Studies on the effects of grade retention have received increasing attention from economists in recent years. Several of these recent studies are characterized by an attempt to address the inherent endogeneity of the retention policy, proposing a causal econometric evaluation that aims to evaluate the potential policy implications of grade retention. In this sense, the data requirements for tackling endogenous selection in retention are considerable. Longitudinal data, which follows the student's academic path, are fundamental in addressing grade repetition accurately. Nevertheless, the data constraints that many countries face force researchers to also propose creative solutions to overcome this restriction (see Choi *et al*., 2018).

The main identification challenge of the effect of grade failure is that latent school outcomes, those that would be observed in the absence of retention, and the propensity to fail a grade are likely to be simultaneously determined (Manacorda, 2012). The endogeneity present in the decision of whether to retain a student is usually related to unobservable characteristics of the students that also affect their future achievement (e.g. ability, motivation and/or maturity). For instance, students with higher ability have a lower probability of being retained and may also obtain higher school grades. Therefore, the relationship between grade retention and school performance is likely to be driven by omitted variable bias (OVB), and studies that ignore it are likely to be influenced by severe selection bias. This implies that the relationship may be negatively (positively) biased for retained (non-retained) students.

Recent studies using more detailed longitudinal datasets are able to apply different identification strategies that substantially improve the comparability between retained and promoted students. The literature investigating the causal effect of retention has made use of a variety of methodologies to overcome selection in retention. The majority of these studies can be categorized into the following approaches for identifying the causal effect:

1. IVs: the IV approach is widely used in estimating causal relationships addressing omitted variable and selection bias problems. Basically, it aims to isolate the exogenous variation in the treatment to obtain unbiased estimates of the (causal) relationship between the outcome and the predictor. It exploits the exogeneous variation induced by a variable, the 'instrument', which introduces randomness into the assignment, thus reproducing the effect of an experiment. The choice of a valid instrument is the key to the effectiveness of this approach. In this regard, the instrument must satisfy two conditions: (1) the IV has to be correlated with the treatment, thus determining the probability of treatment, and (2) it has to be uncorrelated with the dependent variable, which also implies absence of correlation with the error term. That is, the association between the instrument and the outcome variable can only happen through the association between the instrument and the independent variable of interest (Angrist and Lavy, 1999). In practice, the IV approach is implemented by using the so-called two-stage least squares (2SLS) regression. The first stage consists of regressing the dependent variable as a function of the IVs plus a set of exogeneous variables that are also included as covariates in the second stage. Then, the estimated probability of the treatment from the previous model replaces the observed indicator (original treatment variable) as the independent variable in the second stage (Heckman, 1979).
2. DiD: This approach relies on a quasi-experimental design mainly using longitudinal data from a treatment group, which is exogenously exposed to a policy or intervention, and a control group, where the treatment is absent, to estimate a causal effect. Thus, the effect of a specific intervention is estimated by comparing the changes in outcomes over time based on the differences between the average outcomes for the treatment and control groups before and after the intervention. In particular, DiD removes the potential different time-constant unobserved heterogeneity across

groups (e.g. fixed-effects), in the absence of treatment, by using the difference within each group. Then, the trends in the outcome of interest would be the same for both groups. Therefore, DiD removes biases when comparing the treatment and control groups after the intervention that might arise from permanent differences between the groups. Hence, the causal effect of the treatment comes from the difference between those differences (Gertler *et al*., 2016).

3. RDD: This methodology is used as a way of estimating causal relationships where the treatment is determined by what is known as the 'forcing' or 'running' variable. In other words, the participation in the intervention or treatment is determined as the running variable exceeds a known cut-off point, where the distribution of the treatment indicator changes discontinuously. Thus, the comparison between individuals in a small neighbourhood of this cut-off point ensures that both the treated and control groups share similar characteristics but only some of them receive the treatment (sharp discontinuity). Therefore, this approach exploits the fact that individuals included in the control and the treatment groups differ only with respect to the assignment variable. Hence, the mean causal effect of the intervention is estimated by the difference in outcomes between these two groups. In particular, the RDD approach measures a local treatment effect which is only applicable to those individuals within a fairly small range above and below the threshold of the assignment variable (Lee and Lemieux, 2010). Finally, RDD models are often used alongside the IV approach. This is the case when there are control and treatment observations on both sides of the cut-off point (fuzzy discontinuity). Then this cut-off point is used as an instrument in a 2SLS framework (Imbens and Lemieux, 2008).
4. Structural (dynamic) models: This approach relies on an economic model where the structure of decision making is fully incorporated in the specification of the model. It requires a detailed specification that describes the preferences and constraints of the process in order to identify the structural parameters, which implies the necessity to rely on strong assumptions. These models aim to (1) Identify the main structural parameters of the model, (2) provide a clear insight into the underlying mechanisms associated to the observed behaviour and (3) provide reliable policy counterfactuals. We can distinguish between full structural models and semi-structural models, the latter aiming to identify only a subset of parameters and/or mechanisms rather than full counterfactuals. Nowadays, structural models play a key role in understanding economic behaviour and in policy design, complementing reduced-form approaches (Blundell, 2017).
5. Other approaches: We mainly refer to standard approaches that enhance the fulfilment of the necessary assumptions before estimating the causal effect.
    a. Propensity score methods: These models aim to reduce estimation bias by reducing large differences between treated and non-treated groups in their observable characteristics. The propensity score is defined as the conditional probability of being treated given the covariates. Therefore, it is estimated by modelling the distribution of the treatment variable given the observed covariates. Then, individuals in both groups with (nearly) equal propensity scores will tend to have (nearly) the same distributions in their background covariates. Thus, the estimated propensity score is used in order to reduce bias by means of matching, stratification, regression adjustment or combinations of all three (Rosenbaum and Rubin, 1983; Dehejia and Wahba, 2002). However, propensity score methods are based on selection in the observables but tell us nothing about selection in unobservables. This methodology is normally used to ensure a common support region in the baseline characteristics before the application of the methodologies discussed earlier (i.e. DiD).
    b. Hierarchical models: This methodology takes into account the nested nature of the (educational) data considering students clustered in classrooms, schools and districts, thus controlling for the correlation between students' results within each cluster structure. In particular, these models allow parameters at a lower level of aggregation to vary as a function of parameters at the next higher level, helping to mitigate the aggregation bias problem. Moreover, such

models include random intercepts and random coefficients to correctly specify error structures, thus addressing the estimation precision problem associated with nested data structures (Hox, 1995). On the other hand, hierarchical models are not designed to solve the OVS problem. Thus, this approach is used in combination with the methodologies discussed earlier in order to account for the specific nested structure of the data before proceeding to the estimation of causal effects.

4.3 *Distinction Same-Age/Same-Grade*

The analysis of the recent literature employing different methodologies and exploiting a variety of policies to investigate the causal effect of grade retention reveals different results. An important methodological reason why studies might differ in their conclusions, and in the corresponding policy implications of their results, is related to the point at which researchers estimate achievement effects, as students cannot be matched in both their ages and their grade levels. Estimates on the effect of repetition differ depending on whether students are compared at the same age (e.g. retained and promoted students one year after retention) or at the same grade (e.g. when retained students reach the same grade level as promoted students).[1,2]

Same-age comparisons focus on the evaluation of retained students' achievement after retention against that of their promoted peers, who are one grade ahead. This comparison is made with the original age cohort, within the same period of time and ideally on an age-standardized measure of achievement (e.g. data used by Schwerdt *et al*., 2017). On the other hand, same-grade comparisons compare retained to promoted children at the same grade level, once they have been exposed to the same grade-level material. Basically, there are two different options for implementing this comparison: (1) retained students are 'shifted back' a year, thus their performance is assessed one year later than are their cohort promoted peers; (2) retained students are compared to their new younger classmates. Some researchers argue that same-age comparisons are more appropriate as they evaluate the effectiveness of grade retention using the correct counterfactual: the outcome that students would have obtained in the absence of retention (Roderick and Nagaoka, 2005),[3,4] although retained students may be penalized in the comparison as they have not had the opportunity to cover the same material as the promoted students.

Conversely, same-grade comparisons have the advantage that they capture differences in academic achievement between retained and promoted students who have covered the same amount of material, although this attributes maturation (or age) as well as having an additional year of schooling to the estimated effect of retention, thus benefiting retained students (Fruehwirth *et al*., 2016). In short, same-age comparisons are more appropriate when the objective of the study is to measure cognitive development (e.g. whether students learn more by age 15 whether they have repeated or not), and same-grade comparisons are preferable when the researchers or policymakers aim to evaluate grade specific knowledge (e.g. learning acquired by sixth grade).

Knowing the strengths and limitations of each approach, it is important to highlight that, as achievement is typically measured by grade-specific tests, the retention literature tends to deviate from the standard approach for any (quasi-)experimental analysis comparing outcomes measured at the same point in time (i.e. same-age comparisons), thus evaluating outcomes when students have reached the same grade (i.e. same-grade comparisons). Indeed, 34 out of 42 studies in our revision employ same-grade comparisons. In this respect, the studies using same-age comparisons in this review identify effects which are considerably more negative as compared with the findings of other recent studies using same-grade comparisons. Thus, results associated to same-grade studies in this review could be read as the upper-bound of the true effect of grade retention.

5. Grade Retention and Student Outcomes

The papers selected for this revision can be classified according to different dimensions: methodological approach, identification strategy, type of data or even the country under analysis. However, considering that our main objective is to revise the recent literature assessing the causal effect of grade retention, we have decided to organize the results by the outcome of interest evaluated in each of the studies. Section 5.1 assesses the impact of grade retention on test scores. Section 5.2 reviews the effect of grade retention on other educational and labour market outcomes. Finally, Section 5.3 closes with a short analysis of the effect of grade retention when combined with alternative remedial measures. Moreover, in order to summarize the results, Table A2 shows the signs of the effects considering the outcome of interest as well as the educational level where the policy is applied.

5.1 *Impact of Grade Retention on Test Scores*

Most studies have focused on assessing the impact of grade retention on mathematics, reading and science. In this subsection, studies have been divided by levels of education (kindergarten, primary school and secondary school) and, for primary and secondary school, by the duration of the effect.

5.1.1 *Kindergarten*

We found two assessments that analysed the effect of grade retention in the kindergarten level. First, Dong (2010), based on across year comparison using a control function (CF) approach,[5] showed that repeating kindergarten had positive effects on the retained children's later academic performance (i.e. the retained children would do worse in terms of the first and third grade test scores, were they socially promoted). Results also suggest that these effects were diminishing over time. For example, while the positive effect on the retainees' math test scores was still significant up to third grade, the effect on the reading test scores was not, indicating the existence of heterogeneous effects according to competence assessment. Second, Fruehwirth *et al*. (2016), using structural models, found that retention in kindergarten is estimated to lower achievement by 9%, early retention by 14%, and late retention by only 4%, in both reading and math. Moreover, they found that students who are retained experience considerable achievement losses with respect to those not retained. Finally, the effect varies by age, the time the student is retained and unobserved abilities with the lowest ability students generally being hurt the most by retention.

5.1.2 *Primary School*

a. Short-term effects

As for primary school, we found more evidence using different methodologies (as matching methods, DiD, hierarchical models, generalized estimating equations, RDD and IV among others). Therefore, for the sake of clarity, we will classify these as studies that show negative, positive or non-effects.

Using a DiD approach, Bhattacharya (2007) found that repetition did not add any value. Children who repeated a grade were more likely to experience a decrease in test scores than they would have if they had not been retained. Moreover, Xinxin *et al*. (2010) demonstrated that we can reject the hypothesis that grade retention improves school performance in the Chinese language the year immediately after a student is retained. In some cases (e.g. for the students who repeat Grade 2), grade retention is shown to hurt school performance.

García-Pérez *et al*. (2014) estimated that grade retention decreases test scores in math of repeaters by 54 points. The results showed that if a student repeated at the primary level, she would suffer a

causal decrease in her performance, but this situation could be even worse if this student were subjected to a second-grade retention in secondary school. Diris (2017) found that the effect of grade retention in primary school harmed student achievement in math and science across the distribution. Similarly, Roderick and Nagaoka (2005) showed that, among sixth graders, there is evidence that retention is associated with lower achievement growth. Namen (2018) found a negative relationship between the grade retention on the test score obtained by non-retained students.

Other authors identified positive short-run effects of grade retention. Using RDD, Winters and Green (2012), in their study of the state of Florida, showed a substantial positive effect on student achievement in math, reading and science in the years immediately following the treatment which fades away over time. Schwerdt *et al*. (2017) found that third-grade retention substantially improved students' reading and math achievement in the short run as well as reducing the probability of being retained in later grades. Using the same technique, Jacob and Lefgren (2004) found that retention increases achievement for third-grade students and had little effect on math achievement for sixth-grade students.

Hughes *et al*. (2010) showed that a positive association between retention and math scores was significant. Goos *et al*. (2013b) found that first-grade repeaters seemed to outperform their equally at-risk but continuously promoted grade-mates in math and reading fluency during the retention year, but that this benefit seemed very short-lived. In fact, it even seemed that this benefit had already disappeared completely in second grade. Moser *et al*. (2012) found an initial advantage in achievement for students' repeated first-grade scores compared to their promoted peers' first-grade scores but this effect faded away over time. Green and Winters (2007) found that retained students slightly outperformed socially promoted students in reading in the first year after retention, and these gains increased substantially in the second year. Pereira and Reis (2014), studying the case of Portugal, detected a small positive relationship. Lorence (2014) showed that third graders failing the state-mandated reading test who repeated the grade consistently outperformed in later grades the socially promoted children who also failed the third-grade test. D'Haultfoeuille (2010) described that the short-term effect of grade retention seems more likely to be positive. He focused here on the average effect of retention in the fifth grade on test score achievement one year later. Nunes *et al*. (2018) estimated the impact of retention at the fourth grade on the scores obtained on the sixth-grade exams, controlling for the level of ability at the moment of retention. They found a low positive impact of retaining in low-achieving student. Finally, Im *et al*. (2013), studying the effects of retention in Grades 1 to 5 on students' reading and math achievement, found that retained and continuously promoted students did not differ in any of the outcome measures during the year prior to transition (to middle school), nor did they differ in their post-transition trajectories.

b. Long-term effects

For this case, Schwerdt *et al*. (2017) found that the initially positive effects of third−grade retention became slightly negative in years four and five but were statistically insignificant after six years. Later grade retention was in fact less beneficial: students who were retained earlier rather than later might particularly benefit from the policy. Moreover, the effects of retention appeared to be slightly less positive for black students than for whites or Hispanics. Pereira and Reis (2014) showed that the modest initial positive effect of grade retention on academic performance turned negative in the long run. Similarly, Alet *et al*. (2013) estimated a positive effect of grade repetition during the first cycle on test scores in third grade (short-run) and a negative effect on test scores in sixth grade (medium-run) by across year comparison using a simultaneous equations model. These results are similar to those obtained by Winters and Green (2012), who found a short-term positive effect which faded away over time. Finally, Belot and Vandenberghe (2014) showed that an enhanced threat of grade retention after a legal change did not lead to better medium-term outcomes, even among the segments of that population most at risk of grade repetition.

5.1.3 *Secondary School*

a. Short-term effects

As for secondary school, we found some assessments using very different approaches (IV, matching methods, DiD, growth curve analysis and a model of knowledge-capital accumulation). Gary-Bobo *et al*. (2016), by means of across year comparison, found that grade retention on mathematics test scores was positive but small at the end of grade 9 (secondary school) and that grade repetitions have some usefulness for the weakest students. Moreover, grade repetition reduced the probability of access to grade 9 of all student types. Another positive effect, for both mathematics and reading, is found by Mahjoub (2017). In this case, he showed a positive effect of grade repetitions, between 10% and 25% of the test-score gain's standard deviation. Moreover, Ferreira *et al*. (2018) found that students who have been exposed to higher retention rates obtain a higher score on the language test. Finally, for Lamote *et al*. (2014), grade retention had no negative effect on achievement in the short term (year of retention).

b. Long-term effects

Most articles found negative long-run effects for grade retention at the secondary school level. Belot and Vanderberghe (2014), with a same-age comparison, showed that an enhanced threat of grade retention after 2001 did not lead to better medium-term outcomes either in reading or mathematics, even among the segments of that population most at risk of grade repetition. Similarly, Lamote *et al*. (2014) identified strong negative long-run effects on reading performance. Finally, while Cockx *et al*. (2019) showed that grade retention has a neutral effect on the evaluation in the next grade, in the long run, grade retention and its alternatives had adverse effects on schooling outcomes and, more so, for less able pupils.

5.2 *Impact of Grade Retention on Other Educational and Labour Market Outcomes*

i. Impact of grade retention on graduation rates

We found four papers that analysed the causal relationship between grade retention on reaching the secondary school.

In the case of grade retention in primary school, Andrew (2014) found that retaining a child in early primary school reduces their odds of high school completion by about 60% using an across year comparison by a propensity score matching and sibling fixed-effects models. Furthermore, Jacob and Lefgren (2009), using an RDD, showed that retention among younger students does not affect the likelihood of high school completion, but that retaining low-achieving eighth-grade students in elementary school substantially increases the probability that these students will drop out of high school.

As for studies that assess the effect of grade retention in secondary level, Brodaty *et al*. (2014) showed that the elasticity with respect to grade repetition risk (or risk of delay) is negative and very important regarding enrolment in college in France. Cockx *et al*. (2019) found that pupils repeating (for the first time) grade 8 have a 14 percentage point lower chance of graduating from high school. Moreover, Mahjoub (2017) concluded that grade repetition improves the probability of graduating from junior high school by 2.5 probability points using IV and matching estimators.

ii. Impact of grade retention on school dropout

In this subsection, we analyse the impact of grade retention on the dropout effect. For the case where the retention was detected in primary school, Glick and Sahn (2010) showed that repeating students are more likely to leave school before completing primary school than students with similar ability who are not held back, pointing to the need for alternative measures to improve the skills of lagging children. Moreover, Cabrera-Hernández (2016) found that abolishing grade retention would reduce dropout rates

in primary school using two-way fixed effects models in a seven-year panel of schools. Finally, Eide and Showalter (2001) using IVs showed that grade retention may have some benefit to students by lowering dropout rates.

As for retention in secondary school, Cockx *et al*. (2019) found that retention of the lowest ability students decreases their academic achievement and has a significantly positive effect on high school dropout, applying a dynamic discrete choice model. Moreover, Hill (2014) showed that increasing the share of repeaters in a given course results in a moderate, significant increase in the probability of course failure for first-time course-takers. The distributional effects also suggested that course repeaters are more likely to distract classmates who are located in similarly low parts of the achievement distribution rather than high achievers. Finally, Manacorda (2012), using a fuzzy RDD, showed that grade failure leads to dropout and lower educational attainment four to five years after failure in the order of 0.8 school years.

As for the grade retention in different levels, Eren *et al*. (2017) showed that potential grade retention, even at fourth grade, increases the chances that a student will drop out of school at a later point in time. Specifically, the adverse effects of potential retention on the likelihood of dropping out were observed only for fourth grade male students. As for the eighth-grade sample, similar to Jacob and Lefgren (2009), female students seemed to be much more affected by grade retention.

5.2.1 *Impact of Grade Retention on Other Educational Outcomes*

For this case, Goos *et al*. (2013b) found that first-grade repeaters seem to lag behind in several psychosocial skills, for at least a part of their primary school career, in comparison to their similarly at-risk grade-mates who got promoted. Moreover, they found that first-grade repeaters seem less likely to repeat another primary school grade, but at the same time, that they were more likely to transition to a special education primary school, move to another primary school and to be placed in a less demanding track in the first year of secondary education than equally at-risk but promoted first graders. Evaluating grade retention at different educational levels, Ou and Reynolds (2010) concluded that retention is associated significantly with lower rates of participation in postsecondary education. Late retention (between fourth and eighth grades) was linked more strongly to lower rates of postsecondary education than early retention (between first and third grades). Finally, the study by Kretschmann *et al*. (2019) focused on motivational outcomes. In particular, the authors analysed differences in learning and achievement motivation, scholarly interests and academic self-concept between those students being retained and those being promoted with the same age. Results showed a decline in all these outcomes for retained students, even before repetition occurs and with a perdurable effect within the next year after repetition that vanishes two years after retention.

5.2.2 *Impact of Grade Retention on Employment Outcomes*

In this case, Eide and Showalter (2001), using IVs, showed that grade retention may have some benefit to students by raising labour market earnings. Babcock and Bedard (2011) found that grade retention is associated to increases in average male hourly wages. Furthermore, the observed positive wage effect is not limited to the lower tail of the wage distribution but appears to persist throughout the distribution. Finally, Brodaty *et al*. (2013) found that delay has a significant, robust and negative impact on the wages of young workers. Thus, the analysis of the effect of grade retention on labour market outcomes reveals mixed results, documenting negative as well positive estimates. Therefore, before we can draw conclusions it is necessary to better understand these contrasting results as the evidence is still very limited.

5.3 *Impact of Grade Retention Combined with Alternative Remedial Measures*

In this subsection, we analyse evidence from the USA based on studies assessing the effects of retention under the high-stakes testing in Chicago, Florida and Louisiana. Following the impact of the emerging accountability systems aiming to improve academic performance, these studies mostly show positive short-term effects associated to early grade retention. It is important to note that under these policies, grade retention is combined with remedial interventions such as summer school, instructional support and better quality teachers.

In the case of Florida, Greene and Winters (2007), using IVs, showed that retained students slightly outperformed socially promoted students in reading in the first year after retention, and these gains increased substantially in the second year. Moreover, Winter and Green (2012) using an RDD approach with Florida data analysed the causal effect of remediation policy (retention, high-quality teacher and summer school) five years after intervention. Exposure to these interventions had a substantial positive effect on student achievement in math, reading and science in the years immediately following the treatment and dissipates over time. They cannot completely disaggregate the effect of retention from that of summer school or teacher assignments during the retained year. However, they provided some evidence that the policy's requirement that a student be assigned to a high-quality teacher the following year does not appear to drive the effects from treatment.

Schwerdt *et al*. (2017), using 2SLS model regression discontinuity with data from US Florida Department of Education, analysed the combined effect of retention and summer school. Results showed a small positive short-term effect on achievement for third graders but not for sixth graders.

The most recent study using data from county-level school districts in Florida and an RDD approach, by Figlio and Özek (2020), showed that the combination of retention and institutional support in third grade substantially increases the English skills of students. In this case, the instructional support, which will occur during the year immediately after retention, consisted of the placement of the students under high-performing teachers using proven-effective teaching strategies.

Regarding high-stakes testing in Chicago, Jacob and Lefgren (2004) using RDD estimates found that summer school and grade retention increased academic achievement in reading and mathematics and that these positive effects remained substantial at least two years following the completion of the program. Specifically, this result reflected the impact of summer school and grade retention with incentives (students had to pass an August exam to avoid retention). Jacob (2005) showed that the high-stakes testing policy led to substantial increases in reading performance on the high-stakes test. Moreover, Jacob and Lefgren (2009) by means of RDD estimates found that grade retention, combined with a six-week summer school program, had no impact on high school completion for sixth-grade low-achievement students but increased the likelihood of dropping out among younger eighth-grade students.

Roderick and Nagaoka (2005), applying an RDD methodology, showed that grade retention did not proffer any academic benefits to third graders who were retained in Chicago. Moreover, in sixth grade the retention was associated with negative growth in achievement.

Finally, Eren *et al*. (2017) examined the potential effects of summer school and grade retention on high school completion and juvenile crime. Drawing on data from the State Agencies in Louisiana and using the RDD approach, they found that grade retention increased the propensity of a student to drop out of school. In case of eighth grade, remedial education (summer school) provided a positive benefit by decreasing the likelihood that a student might drop out. As for fourth-grade retention, however, they did not find any effect of summer school assignment.

6. Alternative Policies to Grade Retention

Grade repetition is not only an ineffective policy, but also it has a negative impact on academic performance, a rare outcome for educational policies. It is also expensive and has equity implications

as its effects are heterogeneous across subgroups. As has been seen, the application of grade repetition varies widely across countries. Thus, policymakers in countries such as Belgium, Spain or Portugal, where the use of grade repetition is widespread, may be interested in implementing alternative policies. However, mere social promotion does not seem to be a valid alternative either (Darling-Hammond, 1998). In the following lines, we provide a brief non-exhaustive review of different, not exclusive, policies. In this sense, we restrict this section to policies applied in countries where grade repetition is applied infrequently, as all educational policies which increase educational performance may have an impact on the risk of grade repetition. The measures reviewed below are founded on two principles: early intervention and individualized treatment.

The successful application of alternative policies to grade repetition needs teachers to identify those students at risk: classroom assessments may better inform teachers (Dennis *et al*., 2012), preferably at early stages in the educational system. Allensworth and Easton (2007) and Balfanz *et al*. (2009) suggested that students at risk can be identified as early as late primary education. Freeman and Simonsen (2015) reviewed educational practices which may reduce school dropout and school completion rates and suggested that early intervention is probably one of the most efficient ways of achieving these aims. This is coherent with the strand in the literature which demonstrates the lasting effects of early schooling (see Carneiro and Heckman (2004), Cunha and Heckman (2008) or Almond and Currie (2011), for example).

A first set of measures for improving the performance of low-performing children consists in the implementation of effective teacher practices. Among some of the practices which have been proved to have a potential for improving the academic achievement of low-performing students are multi-age grouping (creating combined classes with students from different years) and looping (a teacher or set of teachers remains with the same group for some years). Leuven and Ronning (2016) showed, for the Norwegian case, that mixed-grade classrooms may have a positive impact on academic performance. While causal evidence on looping practices is still scarce, most of the existing analyses, such as Nichols and Nichols (2002), Cistone and Shneyderman (2004) or Franz *et al*. (2010) showed a positive association between looping and academic achievement. All these studies analysed the practice of looping in which a teacher follows the students from one grade to the next, instructing the same core group for at least two school years. The results of the former study are based on multiyear looping students who remained up to four years with the same teacher whereas the latter studies focused on pupils who looped during two years.

A second set of measures consists in increasing instructional time for poor performers –this is, in fact, at the core of grade repetition as an educational policy, and the application of additional tuition. Remedial tutoring is an expensive but effective alternative to grade repetition (Slavin *et al*., 2011), which may be applied within or outside schools, during school hours, during playground time or after school and during the academic semester or during the holiday break. Tutoring can also be performed on a one-to-one or on a group basis, although a key characteristic of this kind of programmes is the individualized treatment of students. As it may be seen, very heterogeneous types of programmes may be labelled as "remedial tutoring". The meta-analysis on tutoring models for improving reading competencies performed by D'Agostino and Murphy (2004), Ritter *et al*. () and Slavin *et al*. (2011) revealed that one-to-one tutoring is more effective than group tutoring and that volunteers are less effective than teachers as tutors. Interestingly, cooperative learning practices, where classmates tutor children with learning difficulties, seem to be highly effective. Some recent examples of effective within-school teacher-led one-to-one programmes are the *Catch Up Numeracy* (Rutt *et al*., 2014), *Catch Up Literacy* (Rutt *et al*., 2015) or *Switch-on Reading* (Godard *et al*., 2014), which were applied in the United Kingdom.

Tutoring can also take place in out-of-school time. For instance, an interesting example of this type of compensatory tutoring is provided by Jacob and Lefgren (2004), who showed the positive impact of a programme consisting of providing remedial classes to children with low socioeconomic status backgrounds during the summer break in Chicago. Similar programmes have been applied during the summer break in Baltimore (Borman and Dowling, 2006) and Iowa (Kim, 2006).

In East Asian countries, such as South Korea or Japan, additional tuition is mainly, although not exclusively, provided by private institutions. East Asian families invest heavily in private tutoring (Bray and Kwo, 2014) and, although this phenomenon may have negative side effects (Bray, 2013), which have led some governments to regulate private tutoring activities – see Choi and Choi (2016) for a description of the struggle of the Korean authorities to regulate the opening hours of *hakwon* (private tutoring institutions) – evidence suggests its positive impact on academic performance (Berberoğlu and Tansel, 2014; Hof, 2014).

Finally, a third option consists in making educational systems and curriculums more flexible, allowing students to catch up with their colleagues or, otherwise, to follow alternative paths. The idea is to remove dead ends and allow low-performing students to have some extra time to acquire basic competencies and skills in key subjects while advancing in other ones in which they performed adequately. However, similar regulations on grade repetition may lead to very different repetition rates (Eurydice, 2011, p. 60). In Europe, automatic promotion is established in very few countries (Iceland and Norway and, although not explicitly regulated, it is also implemented *de facto* in England). However, countries such as Slovenia, Sweden or Finland, where grade repetition is allowed, also have low repetition rates. At the same time, limiting the number of repeated years does not seem to be especially effective. Indeed, most countries with high repetition rates, such as Spain, Belgium or France, have this kind of legal limitations. Moreover, curricular flexibility itself does not guarantee lower repetition rates. For instance, Spanish lower secondary students are allowed to move to the next year if they have failed two, or even three, subjects, which they should pass the next academic year. All in all, this evidence suggests that it might be more relevant to focus efforts on changing parental and teachers' beliefs and providing effective tools to teachers rather than introducing regulatory changes on grade repetition.

7. Discussion and Policy Implications

An important consideration regarding the policy implications associated to grade retention is related to the point at which researchers estimate the effects. A common pattern in the studies surveyed in this review shows that early grade retention has short-term positive effects that disappear as students advance to higher grades, and become small or statistically insignificant after several years. On the other hand, the effects identified are more harmful at later grades during primary school (i.e. grades 3 and 6) and turn into severe negative effects in secondary school (Diris, 2017). Overall, these results suggest that early grade retention may have positive effects on short-term student outcomes, but retention in higher grade levels may be more detrimental. This is consistent with the theoretical literature on childhood investments that highlights the potential benefits of interventions earlier in life (Cunha and Heckman, 2008). Additionally, it is likely that the timing of retention has a different impact on future outcomes due to the heterogeneous effect of the policy across the ability distribution.

The policy implications associated to the estimates of the effect of grade retention also depend on the methodology used by researchers to construct adequate comparison groups of retained and promoted children. Table A1 shows that the recent literature aiming to uncover causal effects frequently employs RDD, exploiting the fact that retention is often based on specific achievement thresholds defining the conditions for promotion (e.g. Jacob and Lefgren, 2004, 2009; Roderick and Nagaoka, 2005; Greene and Winters, 2007, 2012; Manacorda, 2012; Schwerdt *et al*., 2017). Overall, the results of these studies suggest that the causal effect of (early) grade retention for students near the cut-off threshold is either positive or null, although it becomes negative when retention happens at later educational stages (i.e. sixth grade). In any case, even the positive impacts hold only for a very short period time as they seem to fade out after a few years.

The main concern associated to the results of these studies is that they limit the analysis to those students within the narrow region of the cut-off achievement, thus evaluating the impact of the policy

only for this potentially small group of marginally affected students (i.e. those who are on the margin of being retained). Thus, they are not able to make inferences about the effect of the policy on low-performing students, who in turn are often the target of retention policies (Fruehwirth *et al*., 2016). In this respect, Green and Winters (2007) propose a comparison of RDD results with those reported by across-year comparisons looking for differences associated to the non-linear effect of retention policies across the ability distribution. Moreover, drawing inferences from RDD results might not be appropriate for either those institutional settings in which retention does not depend on specified achievement thresholds or countries that make an intensive use of the policy as few students would be close to the achievement threshold.

Considering the importance of the policy implications of the potential heterogeneity of the treatment effects of grade retention, the other most common methodological strategy for dealing with the endogeneity of the effect, IV, does not solve the problem either. The IV estimator is a weighted average of the marginal treatment effects; thus, it might not be correctly identifying the relevant effects in the presence of heterogeneity (Heckman and Vytlacil, 2005; Heckman, 2010). Evidence concerning the heterogeneity in the effects of retention across student characteristics is still limited. In this regard, a handful of recent studies focus on the heterogeneous effects of grade retention mostly by using structural (dynamic) models (e.g. Fruehwirth *et al*., 2016; Gary-Bobo *et al*., 2016; Diris, 2017; Cockx *et al*., 2019). These studies consider that the learning process is highly dynamic and retention is unlikely to affect different types of students equally. The effect of grade retention varies depending on students' abilities, and the timing of retention: (1) lower ability students have a higher probability of being retained and are likely to learn at a slower rate, (2) higher ability students who are retained can better take advantage of opportunities post-retention, (3) students retained at different grades differ in unobservable characteristics, such as abilities, leading to dynamic selection. The results of this literature show that there is substantial heterogeneity in the effects of retention across the achievement distribution, affecting more adversely lower ability students when considering early grade retention. Therefore, it is necessary to improve the understanding of this heterogeneity in the effect of grade retention to better implement the corresponding educational policies.

At this point, it is important to consider to what extent the results are affected by particular institutional settings. Evidence from the USA based on studies assessing the effects of retention under high-stakes testing (e.g. CPS, Florida Policy), following the impact of the emerging accountability systems aiming to improve academic performance, mostly shows positive short-term effects associated to early grade retention. It is important to note that under these policies, grade retention is combined with remedial interventions: retained students are given the opportunity to attend summer school programs prior to the next school year as well as being assigned to 'high performing' teachers during their retained year. Thus, this is likely to attenuate the negative consequences of the retention and reduce the probability of failing in the near future (Manacorda, 2012). Therefore, in terms of policy implications, estimates associated to the impact of the policy should be understood as the combined effects of retention and these additional measures.

Finally, further attention should be paid to the long-term effects and potential externalities of grade retention policies. There are only a few studies that focus on the evaluation of the effects of repetition in the long term. The first contribution is by Eide and Showalter (2001) who studied the impact of grade repetitions on wages in the USA. They find that grade repetitions have a positive but non-significant effect on wages.[6] Brodaty *et al*. (2013) showed that, conditional on individuals' highest credential, grade repeaters tend to obtain lower wages in the labour market. Thus, retention may lead to lower wages because of the delayed entry into the labour market, and also because retention can be a negative signal to employers leading to worse labour market conditions. Fruehwirth *et al*. (2016) also noted that the negative consequences of the year lost through retention in terms of additional schooling and wages could easily outweigh any short-term positive benefits of the policy. Finally, Gary-Bobo *et al*. (2016) highlighted the importance of considering the (potential) substantial cost of grade retention, in terms of

delayed and lower wages, when evaluating this educational policy. In short, although the evidence is still very scarce, these recent studies indicate that the long-term effects of grade retention seem to be detrimental.

At the same time, there are reasons to believe that there are potential implications of retention policies for all students, retained and non-retained. Previous work in applied economic theory predicted that retention policies influenced effort, expectations and practices of students and their teachers, therefore influencing the outcomes of retained as well as promoted students (see literature review in Babcock and Bedard, 2011). In particular, the study by Hill (2014) evaluated the negative externalities of grade retention in high school. The results showed that increasing the share of repeaters in a course leads to a significant increase in the probability of failure associated to students taking the course for the first time, affecting especially low-ability students. Similarly, the results in Lavy *et al*. (2012) showed peer effects of retention as the proportion of low-ability peers (most likely having repeated kindergarten or first grade) is negatively correlated with the academic achievement of regular students. These results suggest that the negative externalities brought into play by repeaters emerge because they are both low-achieving and repeating students. This is also important as we have seen in this review that policies aimed at reducing grade repetition might not deal with low-achievers but only with students at the margin of being retained. Thus, these studies provide evidence pointing to the fact that grade retention policies may be costly both to the repeating students and their classmates.

All in all, the evidence presented in this study, which aims to unveil causal effects and serve as a guideline to educational policy decision makers, suggests that grade retention in medium to higher course grades may have detrimental effects on future student outcomes, whereas early retention may be more beneficial at best. These positive effects arise especially in specific institutional contexts involving summer schools, instructional support and better quality teachers alongside grade retention. Moreover, the positive effects hold in the short run but they tend to vanish after a few years. On the other hand, the vast cost of grade retention in terms of public resources, the potential implications of the policy associated to its negative externalities combined with the limited evidence on the long-term effects of the policy that indicate the existence of substantial costs related to negative labour market conditions, would make it hard to justify the policy(ies) of grade retention, from a policy-making perspective, without considering alternative measures aimed at helping students at risk, especially those with a low-ability profile.

8. Conclusions

The effects of grade retention are not easy to estimate. This is basically due to the endogenous character of the decision to hold a student back, as students are not randomly selected for grade retention. Thus, most of the early contributions on the topic failed to account for this unobservable heterogeneity, so that drawing inferences from this body of research would be difficult as well as imprecise. Therefore, its reliability for policy recommendations is certainly limited.

Studies that reliably address the endogeneity of the policy have arisen only recently alongside the availability of detailed longitudinal datasets that allow researchers to apply different identification strategies to establish the causal effect of retention. Therefore, based on a systematic literature review, this paper provides a comprehensive description of recent empirical studies that do address endogeneity by means of applying a variety of methodologies aiming to estimate the causal effect of grade retention. We believe that this research is potentially useful for policymakers, professionals, researchers and parents interested in knowing whether retention is an effective educational strategy, and which are the consequences of the policy for students, schools and society alike.

Our literature review reveals that rather than considering retention as a binary event, that is, whether a student does or does not repeat a grade, different kinds of retention should be evaluated. As posed by Allen *et al*. (2009), the question of 'what is the effect of grade retention on achievement?' is probably

too broad to guide educational policies. This fact is extremely important as the effects of this educational policy differ systematically regarding the timing of the implementation associated to the policy (i.e. early vs. late retention), on the basis of the comparison group used (age vs. grade comparisons), the particular institutional settings under which the policy is carried out (US high stakes tests vs. European countries) and the post-retention years considered (short- vs. long-run effects). Furthermore, recent literature highlights the importance of the dynamic character of retention policies and their relevance for the effectiveness of the policy. Students who are promoted, having a similar likelihood of being retained as compared with their peers affected by the policy, might have a greater probability of being retained or at least of having difficulties during the course of the next few years, thus facing the possibility of being described as a 'delayed intervention' group. Thus, the omission of these delayed effects and subsequent performance and interventions in the evaluation of the policy is potentially important (Moser *et al*., 2012). Finally, it is also necessary to include in the analysis the potential externalities of grade retention affecting both retained and non-retained students.

According to the results of our systematic review, the benefits of grade retention policies are observed only, and not always, in the short term and are mostly associated to very specific institutional environments. On the other hand, the policy is clearly associated to relevant costs in terms of academic outcomes, career choice, delayed labour market participation, forgone income, formation of undesirable traits as well as the necessity of using a vast amount of public resources to implement it. Even considering the potential positive threat/motivational effects, which do not imply that the policy is directly beneficial to those who are retained, the policy would not be justified from a policy-making perspective. Hence, we conclude that grade retention is unlikely to be an efficient public policy: its impact on student performance, when positive, is weak, and the negative consequences can easily outweigh its potential benefits. It is therefore necessary to consider alternative policies to grade retention, or policies to be used in combination with it, in order to enhance the achievement of low performers, especially those students characterized by low-ability traits.

Finally, we would like to stress that most of the studies included in this survey are possible due to the availability of adequate data. It is crucial for researchers who aim to convincingly evaluate the effectiveness of educational policies in the long term to have access to appropriate longitudinal datasets at the student or school level. It is necessary then that institutions and policymakers make an effort to build and ensure access to this type of relevant information increasing the quantity and quality of the policy evaluation of particular interventions.

Acknowledgments

The authors gratefully acknowledge the financial support from the Spanish Ministry of Economy and Competitiveness (project EDU2016-76414-R) and from the Areces Foundation through its XIII National Contest for Research in Social Sciences. We are indebted to Dr Andrés Fernández Ramos for his invaluable guidance on using bibliographic databases. Javier Valbuena would like to thank the support from the project S23_20R (Grupo de Investigación de Referencia en Economía Pública, Gobierno de Aragón). Mauro Mediavilla is grateful for the support from Regional Government of Valencian Community (Conselleria d'Hisenda i Model Econòmic; project HIECPU/2019/2).

Notes

1. Allen *et al*. (2009) already found that retention effects were less negative when same-grade comparisons were employed.
2. It is important to notice that the estimation of the treatment effect for retained students implies the comparison of the outcome variable of those that were retained with the outcome variable if they

were not retained, conditioned on the possibility of retention being present in the actual and the counterfactual scenarios.

3. Using same-grade comparisons would evaluate the benefit of adding an extra year of education through retention, which should be compared to other alternative policies that increase instruction time (i.e. investing in preschool or using transitional years). Thus, from an economic point of view, same-grade comparisons should be evaluated against the cost of educational expenditures associated to retention and the opportunity cost of losing a year (Diris, 2017).
4. This would correspond to estimating the causal effect of grade retention polices, thus comparing the outcome variable in a counterfactual scenario that mimics the absence of retention in an education system.
5. The CF approach is essentially an IVs method. In particular, the structural equation of interest contains at least one explanatory variable that is endogenous. Moreover, the exogenous variation explaining the endogenous variables is induced by IVs, which provide a separate variation in the residuals obtained from a reduced form, and these residuals serve as the CFs. Thus, by adding appropriate CFs, usually estimated in a first-stage, the endogenous explanatory variables become appropriately exogenous in a second-stage estimating equation. Hence, they provide a consistent estimation of the causal effect of a policy intervention. The CF approach can be applied to various linear and nonlinear models (Wooldridge, 2015).
6. Furthermore, recent literature has noted that the effect of grade retention is imprecisely estimated (Brodaty *et al*., 2013).

References

Agasisti, T. and Cordero, J. (2017) The determinants of repetition rates in Europe: Early skills or subsequent parents' help? *Journal of Policy Modeling* 39(1): 129–146.

Alet, E., Bonnal, L. and Favard, P. (2013) Repetition: medicine for a short-run remission. *Annals of Economics and Statistics* 111/112: 227–250.

Allen, C.S., Chen, Q., Willson, V.L. and 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.

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*. Chicago, IL: Consortium on Chicago School Research.

Almond, D. and Currie, J. (2011) Human capital development before age five. In O. Ashenfelter & D. Card (eds.), *Handbook of Labor Economics* (Vol. 4B, pp. 1315–1486). Amsterdam: Elsevier.

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.

Angrist, J.D. and Lavy, V. (1999) Using maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics* 114(2): 533–575.

Babcock, P. and Bedard, K. (2011) The wages of failure: New evidence on school retention and long-run outcomes. *Education Finance and Policy* 6(3): 293–322.

Balfanz, R., Fox, J.H., Bridgeland, J.M. and McNaught, M. (2009) *Grad Nation: A Guidebook to Help Communities Tackle the Dropout Crisis*. Washington, DC: America's Promise Alliance.

Bali, V., Anagnostopoulos, D. and Roberts, R. (2005) Toward a political explanation of grade retention. *Educational Evaluation and Policy Analysis* 27(2): 133–155.

Bedard, K. and Dhuey, E. (2006) The persistence of early childhood maturity: International evidence of long-run age effects. *Quarterly Journal of Economics* 121: 1437–1472.

Belot, M. and Vandenberghe, V. (2014) Evaluating the 'threat' effects of grade repetition: Exploiting the 2001 reform by the French-Speaking Community of Belgium. *Education Economics* 22(1): 73–89.

Berberoğlu, G. and Tansel, A. (2014) Does private tutoring increase students' academic performance? Evidence from Turkey. *International Review of Education* 60(5): 683–701.

Bhattacharya, S. (2007) The Effect of Grade Retention on Child Test Scores. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1024090. Date: 05/12/2020.

Blundell, R. (2017) What have we learned from structural models? *American Economic Review* 107(5): 287–292.

Bonvin, P., Bless, G. and Schüpbach, M. (2008) Grade retention: Decision-making and effects on learning as well as social and emotional development. *School Effectiveness and School Improvement* 19(1): 1–19.

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.

Borodankova, O. and Coutinho, A.S.A. (2011) *Grade Retention during Compulsory Education in Europe: Regulations and Statistics*. Education, Audiovisual and Culture Executive Agency, European Commission. Available from EU Bookshop.

Bray, M. (2013) Shadow education: Comparative perspectives on the expansion and implications of private supplementary tutoring. *Procedia - Social and Behavioral Sciences* 77: 412–420.

Bray, M. and Kwo, O. (2014) *Regulating Private Tutoring for Public Good: Policy Options for Supplementary Education in Asia*. Hong Kong: UNESCO and Comparative Education Research Centre.

Brodaty, T., Gary-Bobo, R.J. and Prieto, A. (2013) *Does speed signal ability? The impact of grade retention on wages*. CEPR Discussion Papers. London, UK: Centre for Economic Policy Research.

Brodaty, T., Gary-Bobo, R.J. and Prieto, A. (2014) Do risk aversion and wages explain educational choices? *Journal of Public Economics* 117: 125–148.

Cabrera-Hernández, F. (2016) *Leave them kids alone! The positive effect of abolishing grade retention on pupils' dropout rates: Evidence from a policy change*. Working Paper, Department of Economics University of Sussex.

Carabaña (2015) Repetir hasta 4° de Primaria: determinantes cognitivos y sociales según PIRLS. *Revista de la Asociación de Sociología de la Educación* 8(1): 7–27.

Carneiro, P. and Heckman, J. (2004) Human capital policy. In J. Heckman & A. Krueger (eds.), *Inequality in America: What Role for Human Capital Policies* (pp. 77–240). Cambridge, MA: MIT Press.

Center for Mental Health in Schools at UCLA. ((2008) Update) Grade Retention: What's the Prevailing Policy and What Needs to be Done? A Center Policy & Practice Analysis Brief, Los Angeles, CA.

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: 544–559.

Choi, H. and Choi, Á. (2016) Regulating private tutoring consumption in Korea: Lessons from another failure. *International Journal of Educational Development* 49: 144–156.

Choi, Á., Gil, M., Mediavilla, M. and Valbuena, J. (2018) The evolution of educational inequalities in Spain: Dynamic evidence from repeated cross-sections. *Social Indicators Research* 138(3): 853–872.

Cistone, P. and Shneyderman, A. (2004) Looping: An empirical evaluation. *International Journal of Educational Policy, Research, and Practice: Reconceptualizing Childhood Studies* 5(1): 47–61.

Cockx, B., Picchio, M. and Baert, S. (2019) Modeling the effects of grade retention in high school. *Journal of Applied Econometrics* 34(3): 403–424.

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

Corman, H. (2003) the effects of state policies, individual characteristics, family characteristics and neighbourhood characteristics on grade repetition in the United States. *Economics of Education Review* 22(4): 409–420.

Crothers, L.M., Schreiber, J.B., Schmitt, A.J., Bell, G.R., Blasik, J., Comstock, L.A., Greisler, M.J., Keener, D. and King, J.M. (2010) A preliminary study of bully and victim behavior in old-for-grade students: Another potential hidden cost of grade retention or delayed school entry. *Journal of Applied School Psychology* 26: 327–338.

Cunha, F. and Heckman, J. (2008) Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources* 43: 738–782.

D'Agostino, J.V. and Murphy, J.A. (2004) A meta-analysis of reading recovery in United States schools. *Educational Evaluation and Policy Analysis* 26(1): 23–28

Darling-Hammond, L. (1998) Avoiding both grade retention and social promotion. *The Education Digest* 64(3): 48–53.
Dehejia, R.H. and Wahba, S. (2002) Propensity score-matching methods for nonexperimental causal studies. *The Review of Economics and Statistics* 84: 151–161.
Dennis, D.V., Kroeger, D.C., O'Byrne, W.I., Meyer, C.K., Kletzein, S.B., Huddleston, A.P. and Gilrane, C. (2012) Test-based grade retention. *Education Policy Analysis Archives* 22: 1–31.
De Witte, K., Nicaise, I., Lavrijsen, J., Van Landeghem, G., Lamote, C. and Van Damme, J. (2013) The impact of institutional context, education and labour market policies on early school leaving: A comparative analysis of EU countries. *European Journal of Education* 48(3): 331–345.
d'Haultfoeuille, X. (2010) A new instrumental method for dealing with endogenous selection. *Journal of Econometrics* 154(1): 1–15.
Diris, R. (2017) Don't hold back? The effect of grade retention on student achievement. *Education Finance and Policy* 12(3): 312–341.
Dong, Y. (2010) Kept back to get ahead? Kindergarten retention and academic performance. *European Economic Review* 54(2): 219–236.
Eide, E.R. and Showalter, M.H. (2001) The effect of grade retention on educational and labor market outcomes. *Economics of Education Review* 20(6): 563–576.
Elder, T.E. and Lubotsky, D.H. (2009) Kindergarten entrance age and children's achievement impacts of state policies, family background, and peers. *Journal of Human Resources* 44(3): 641–683.
Eren, O., Depew, B. and Barnes, S. (2017) Test-based promotion policies, dropping out, and juvenile crime. *Journal of Public Economics* 153: 9–31.
Eurydice (2011) *Grade Retention During Compulsory Education in Europe: Regulation and Statistics*. Brussels: European Comission. https://doi.org/10.2797/50570.
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): 327–341.
Ferreira, M., Golsteyn, B. and Parra-Cely, S. (2018) The effect of grade retention on secondary school performance: Evidence from a natural experiment. IZA Discussion Paper No. 11604.
Figlio, D. and Özek, U. (2020) An extra year to learn English? Early grade retention and the human capital development of English learners. *Journal of Public Economics* 186: 104–184.
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.
Freeman, J. and Simonsen, B. (2015) Examining the impact of policy and practice interventions on high school dropout and school completion rates: A systematic review of the literature. *Review of Educational Research* 85(2): 205–248.
Frey, N. (2005) Retention, social promotion and academic redshirting: What do we know and need to know? *Remedial and Special Education* 26(6): 332–346.
Fruehwirth, J.C., Navarro, S. and Takahashi, Y. (2016) How the timing of grade retention affects outcomes: Identification and estimation of time-varying treatment effects. *Journal of Labor Economics* 34(4): 979–1021.
García-Pérez, J.I., Hidalgo-Hidalgo, M. and Robles-Zurita, J.A. (2014) Does grade retention affect students' achievement? Some evidence from Spain. *Applied Economics* 46(12): 1373–1392.
Gary-Bobo, R.J., Goussé, M. and Robin, J.M. (2016) Grade retention and unobserved heterogeneity. *Quantitative Economics* 7: 781–820.
Gertler, P.J., Martinez, S., Premand, P., Rawlings, L.B. and Vermeersch, C.M. (2016) *Impact Evaluation in Practice*, 2nd ed. Washington, DC: World Bank Publications.
Glick, P. and 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.
Gomes-Neto, J.B. and Hanushek, E.A. (1994) Causes and consequences of grade repetition: Evidence from Brazil. *Economic Development and Cultural Change* 43(1): 117–148.

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

Goos, M., Van Damme, J., Onghena, P., Petry, K. and de Bilde, J. (2013b) 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.

Gorard, S., See, B.H. and Siddiqui, N. (2014) *Switch-On Reading. Evaluation Report and Executive Summary*. London: Education Endowment Foundation.

Greene, J.P. and Winters, M.A. (2007) Revisiting grade retention: An evaluation of Florida's test-based promotion policy. *Education Finance and Policy* 2(4): 319–340.

Guèvremont, A., Roos, N.P. and Brownell, M. (2007) Predictors and consequences of grade retention: Examining data from Manitoba. *Canada. Canadian Journal of School Psychology* 22(1): 50–67.

Haney, W. (2000) The myth of the Texas miracle in education. *Education Policy Analysis Archives* 8: 41.

Heckman, J.J. (1979) Sample selection bias as a specification error. *Econometrica* 47: 153–161.

Heckman, J.J. (2010) Building bridges between structural and program evaluation approaches to evaluating policy. *Journal of Economic Literature* 48(2): 356–398.

Heckman, J.J. and Vytlacil, E. (2005) Structural equations, treatment effects, and econometric policy evaluation. *Econometrica* 73(3): 669–738.

Hill, A.J. (2014) The costs of failure: Negative externalities in high school course repetition. *Economics of Education Review* 43: 91–105.

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

Holmes, C.T. and Matthews, K.M. (1984) The effects of non-promotion on elementary and junior high school pupils: A meta-analysis. *Review of Educational Research* 54(2): 225–236.

Hof, S. (2014) Does private tutoring work? The effectiveness of private tutoring: A nonparametric bounds analysis. *Education Economics* 22(4): 347–366.

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

Hughes, J.N., Chen, Q., Thoemmes, F. and Kwok, O. (2010) An investigation of the relationship between retention in first grade and performance on high stakes tests in third grade. *Educational Evaluation and Policy Analysis* 32(2): 166–182.

Ikeda, M. and García, M. (2014) Grade repetition: A comparative study of academic and non-academic consequences. *OECD Journal: Economic Studies* 2013(1): 269–315.

Im, M.H., Hughes, J.N., Kwok, O., Puckett, S. and Cerda, C.A. (2013) Effect of retention in elementary grades on transition to middle school. *Journal of School Psychology* 51(3): 349–365.

Imbens, G. and Lemieux, T. (2008) Regression discontinuity designs: A guide to practice. *Journal of Econometrics* 142(2): 615–635.

Jacob, B. (2005) Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools. *Journal of Public Economics* 89(5-6): 761–796.

Jacob, B. and Lefgren, L. (2004) Remedial education and student achievement: A regression-discontinuity analysis. *The Review of Economics and Statistics* 86(1): 226–244.

Jacob, B. and Lefgren, L. (2009) The effect of grade retention on high school completion. *American Economic Journal: Applied Economics* 1(3): 33–58.

Jimerson, S. (2001) Meta-analysis of grade retention research: implications for practice in the twenty-first century. *School Psychology Review* 30(3): 420–437.

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

Jimerson, S. and Ferguson, P. (2007) A longitudinal study of grade retention: Academic and behavioral outcomes of retained students through adolescence. *School Psychology Quarterly* 22(3): 314–339.

Kim, J.S. (2006) Effects of a voluntary summer reading intervention on reading achievement: Results from a randomized field trial. *Educational Evaluation and Policy Analysis* 28(4): 335–355

Kloosterman, R. and de Graaf, P.M. (2010) Non-promotion or enrolment in a lower track? The influence of social background on choices in secondary education for three cohorts of Dutch pupils. *Oxford Review of Education* 36(3): 363–384.

Kretschmann, J., Vock, M., Lüdtke, O., Jansen, M. and Gronostaj, A. (2019) Effects of grade retention on students' motivation: A longitudinal study over 3 years of secondary school. *Journal of Educational Psychology* 111(8): 1432–1446.

Lamote, C., Pinxten, M., Van Den Noortgate, W. and 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(5): 496–514.

Lavy, V., Paserman, M. and Schlosser, A. (2012) Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom. *The Economic Journal* 122: 208–237.

Lee, D.S. and Lemieux, T. (2010) Regression discontinuity designs in economics. *Journal of Economic Literature* 48(2): 281–355

Leuven, E. and Ronning, M. (2016) Classroom grade composition and pupil achievement. *The Economic Journal* 126(593): 1164–1192.

Liddell, C. and Gordon, R. (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: 413–428.

Lorence, J. (2014) Third-grade retention and reading achievement in Texas: A nine year panel study. *Social Science Research* 48: 1–19.

Mahjoub, M.B. (2017) The treatment effect of grade repetitions. *Education Economics* 25(4): 418–432.

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

Mancebón, M.J., Calero, J., Choi, Á. and Ximénez-de-Embún, D.. (2012) The efficiency of public and publicly subsidized high schools in Spain: Evidence from PISA-2006. *Journal of Operational Research Society* 63: 1516–1533.

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

Moser, S.E., West, S.G. and Hughes, J.N. (2012) Trajectories of math and reading achievement in low-achieving children in elementary school: Effects of early and later retention in grade. *Journal of Educational Psychology* 104(3): 603–621.

Namen, O. (2018) The peer effects of grade retention. Working paper, Department of Economics, Universidad del Rosario and Innovations for Poverty Action (IPA).

Neal, D. and Schanzenbach, D.W. (2010) Left behind by design: Proficiency counts and test-based accountability. *Review of Economics and Statistics* 92(2): 263–283.

Nichols, J.D. and Nichols, G.W. (2002) The impact of looping and non-looping classroom environments on parental attitudes. *Educational Research Quarterly* 26(1): 23–40.

Nunes, L.C., Balcão Reis, A. and Seabra, C. (2018) Is retention beneficial to low-achieving students? Evidence from Portugal. *Applied Economics* 50(40): 4306–4317.

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

OECD (2016) *PISA 2015 Results (Volume II): Policies and Practices for Successful Schools, PISA*, Paris: OECD Publishing.

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): 820–834.

Pereira, M. and Reis, H. (2014) *Grade retention during basic education in Portugal: Determinants and impact on student achievement. Economic Bulletin and Financial Stability Report Articles*, Banco de Portugal.

Pérez, V., Rodríguez, J. and Fernández, J. (2009) *Educación y Familia. Los padres ante la educación general de sus hijos en España. Estudios de la Fundación: Serie Economía y Sociedad*. Madrid: Fundación de las Cajas de Ahorros.

Pérez-Díaz, V., Rodríguez, J. and Sánchez, L. (2001) La familia española ante la educación de sus hijos. Colección Estudios Sociales N°5. Barcelona: Fundación "La Caixa".

Ritter, G.W., Barnett, J.H., Denny, G.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(1): 3–38.
Roderick, M. and Nagaoka, J. (2005) Retention under Chicago's high-stakes testing program: Helpful, harmful, or harmless? *Educational Evaluation and Policy Analysis* 27(4): 309–340.
Rosenbaum, P.R. and Rubin, D.B. (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1): 41–55.
Rutt, S., Easton, C. and Stacey, O. (2014) *Catch Up Numeracy: Evaluation Report and Executive Summary*. London, Education Endowment Foundation.
Rutt, S., Kettlewell, K. and Bernardinelli, D. (2015) *Catch Up Literacy: Evaluation Report and Executive Summary*. London, National Foundation for Educational Research.
Schwerdt, G., West, M.R. and Winters, M.A. (2017) The effects of test-based retention on student outcomes over time: Regression discontinuity evidence from Florida. *Journal of Public Economics* 152: 154–169.
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.
Sprietsma, M. (2010) 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: 1–32.
Spruyt, B. (2008) Non-collective educational choices as mechanisms of social reproduction of educational inequality. *Steunpunt Studieen Schoolloopbanen* Louvain.
Tillman, K.H., Guo, G. and Harris, K.M. (2006) Grade retention among immigrant children. *Social Science Research* 35(1): 129–156.
Verachtert, P., De Fraine, B., Onghena, P. and Ghesquière, P. (2010) Season of birth and school success in the early years of primary education. *Oxford Review of Education* 36(3): 285–306.
Werner, E.E. (2013) What can we learn about resilience from large-scale longitudinal studies? In Goldstein S., Brooks R. (eds.), *Handbook of Resilience in Children* (pp. 87–102). Boston, MA: Springer. https://doi.org/10.1007/978-1-4614-3661-4_6.
West, M.R. (2012) Is retaining students in the early grades self-defeating? CCF Brief# 49, Center on Children and Families at Brookings.
Willson, V. and Hughes, J. (2009) Who is retained in first grade? A psychosocial perspective. *Elementary School Journal* 109(3): 251–266.
Winters, M.A. and Greene, J.P. (2012) The medium-run effects of Florida's test-based promotion policy. *Education Finance and Policy* 7(3): 305–330.
Wooldridge, J.M. (2015) Control function methods in applied econometrics. *Journal of Human Resources* 50(2): 420–445.
Wynn, J.L. (2010) *A study of selected teachers' perceptions of grade retention in a Florida school district*. University of South Florida.
Xia, N. and Kirby, S.N. (2009) *Retaining Students in Grade: A Literature Review of the Effects of Retention on Students' Academic and Non-academic Outcomes*. Santa Monica, CA: RAND Corporation.
Xinxin, C., Chengfang, L., Linxiu, Z., Yaojiang, S. 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(6): 544–559.

Appendix

Table A1. Studies Estimating the Causal Effect of Grade Retention on Academic and Employment Outcomes.

Longitudinal databases

Author (year)	Country and database	Methodology	Main results
Alet *et al*. (2013) in *Annals of Economics and Statistics*	France. French Education Ministry database. Panel data. $N = 6700$	Simultaneous equations model that takes into account the hierarchical structure of school systems. Multi-stage econometric model in which identification is ensured through exclusion restrictions as in an IV framework. Same-grade comparisons.	Results showed that early grade repetition (grade 1 or 2) leads to a modest increase in test scores in the very short run (in grade 3) but this positive effect is only transitory as it becomes negative 3 or 4 years after the retention (in grade 6). The treatment corresponds to repetitions during the first cycle and the outcome variable corresponds to the score obtained on thord- or the sixth-grade tests. Primary school.
Andrew, M. (2014) in Social Forces	US. National Longitudinal Survey of Youth 1979, Child and Young Adult supplements (NLSY79-C, YA) and the National Education Longitudinal Study 1988 (NELS88). $N = 8808$	Propensity score matching and sibling fixed-effects models. Mediation analysis. Same-grade comparisons.	The study showed that primary-grade retention has lasting effects on educational attainments well after a student is initially retained. Retaining a child in early primary school reduces her odds of high school completion by about 60% in propensity score matching and sibling fixed-effects models. Models suggest that early primary grade retention scars the educational career mainly at high school completion, though there are important, unconditional effects on college entry and completion as a result. Grade retention in primary school leaves lasting scars on students' educational careers, lowering the odds of completing a high school credential. Primary school.

(*Continued*)

Table A1. *Continued.*

Longitudinal databases			
Author (year)	**Country and database**	**Methodology**	**Main results**
Babcock, P. and Bedard, K. (2011) in *Education Finance and Policy*	US. 1960, 1970 and 1980 Public Use Microdata. 1960–1980 censuses and wage data from the 2000 census and the 2001−2007 ACSs. $N >$ 500,000	Panel data Econometrics: within-state variation of retention rates on wages. The retention rate over grades 1 and 2 is calculated from the observed age and grade outcomes of birth cohorts covering three years. Thus, adult respondents in three-year birth cohorts are mapped to the retention rate associated with the year (and birth state) in which they would have been in first or second grade. Same-grade comparisons.	This analysis offers what may be the first estimates of average long-run impacts of retention on all students. Using within-state variation in primary school retention rates from 1960 to 1980, we find that a 1 standard deviation increase in early grade retention is associated with a 0.7% increase in mean male hourly wages. Further, the observed positive wage effect is not limited to the lower tail of the wage distribution but appears to persist throughout the distribution. Primary school.
Bhattacharya, S. (2007) in SSRN. Rochester	US. 1979 National Longitudinal Survey of Youth (NLSY79) and the NLSY79 Child Survey. Math and reading scores. 1979 to 2002. $N = 4759$	Difference-in-difference propensity score matching estimator aiming to correct for selection bias. Same-grade comparisons.	Results showed that grade repetition, on average, does not add any value in terms of improved mathematics and reading test scores for the repeaters. Children who repeat a grade are more likely to experience a decrease in test scores than they would have if they had not been retained. Retention in Primary school. Outcome: achievement between five and fourteen.
Brodaty *et al*. (2013) in Unpublished manuscript, CREST-ENSAE, France.	France. Génération 92, a large-scale survey conducted by CEREQ. $N = 12{,}310$	IV estimation of the impact of grade repetitions on wages. Same-grade comparisons.	Delay is the part of school-leaving age that is not explained by the highest degree. Variability in delay is mainly due to grade retention. Making use of various instruments, we find a robust, significant and negative impact of delay on wages. A year of delay causes a decrease of the students' beginning- of-career wage around 9%, while at the same time, returns to education are positive with values also around 9%. Primary, secondary and higher education retention.

(*Continued*)

Table A1. *Continued.*

Longitudinal databases

Author (year)	Country and database	Methodology	Main results
Brodaty *et al*. (2014) in *Journal of Public Economics*	France. Génération 92, a large-scale survey conducted by CEREQ. $N = 12{,}500$	Maximum likelihood over a theoretical model in which a student's investment in education maximizes expected utility conditional on public and private information. Same-grade comparisons.	A substantial part of the variance of school-leaving age, conditional on education level or degrees happens to be due to grade retention. Simulations show a strong impact of changes in the probability of grade retention on educational achievement. The elasticity with respect to grade-repetition risk (or risk of delay) is negative and very important regarding enrolment in college. Secondary school.
Cabrera-Hernandez, F. (2016) in WP Department of Economics University of Sussex January	México. Annual national Statistics 911 from the academic year 2006–2007 to 2013–2014. $N = 500{,}000$ schools	Panel data econometrics: two-way fixed effects models. Same-grade comparisons.	This paper evaluates the impacts of an exogenous policy change in Mexico which eliminates retention in-grade for all first- to third-grade students. Estimations show an average reduction in dropout rates after reform implementation. Further findings suggest that eliminating the threat of grade repetition did not affect pupil's performance in standardized tests. Primary school.
Cockx *et al*. (2019) in *Journal of Applied Economics*	Flanders. Two random samples of respondents, one born in 1978 and the other in 1980. $N = 3933$.	Factor analytic dynamic models (FADM): A dynamic discrete choice model Same-age comparisons.	Even if the results indicate that grade retention leads to neutral effects on academic achievement in the short run, in the long run grade retention has adverse effects, because it leads to higher dropout rates, substantial schooling delay and downgrading within the hierarchical tracking system in Flemish high school. Secondary school

(*Continued*)

Table A1. *Continued.*

Longitudinal databases			
Author (year)	**Country and database**	**Methodology**	**Main results**
Fruehwirth *et al*. (2016) in *Journal of Labor Economics*	US. Early Childhood Longitudinal Study of Kindergartners ECLS-K data. $N = 7832$ and $N = 2106$	Structural model: Factor-analytic model allowing for time-varying treatment effects. Same-grade comparisons.	The research shows how the effect of grade retention varies by abilities, by timing of retention, and as time since retention elapses. It finds that students who are retained in kindergarten would have performed as much as 27% higher in the next year if they had not been retained. The paper also finds that the initial losses to achievement diminish over time. By the end of our data, when students are approximately age 11, eliminating grade retention raises achievement by as much as 7% for students who were retained in prior years. This means that these retained students learn 7% less by age 11 than they would have learned if they had not been retained. Kindergarten school.
d'Haultfoeuille, X. (2010) in *Journal of Econometrics*	France. Panel of the French Ministry of Education. $N = 7{,}175$	IV approach Same-grade comparisons.	The short-term effect of grade retention seems more likely to be positive. Primary school.
Diris, R. (2017) in Education Finance and Policy	Some European countries. PISA: multiple waves	Instrumental variable (IV) model Same-age comparisons.	The study evaluates the effect of age-based retention at different stages of education on school achievement (Math, Reading; Science). Grade retention in primary school harms student achievement across the distribution, while delayed school entry can produce positive results for those at the lower end. Primary school

(*Continued*)

Table A1. *Continued.*

Longitudinal databases

Author (year)	Country and database	Methodology	Main results
Dong, Y. (2010) in *European Economic Review*	US. US Early Childhood Longitudinal Study—Kindergarten Cohort 1998−1999 (ECLS-K) by the US National Center for Education Statistics (NCES). $N = 8{,}672$	A control function approach is developed to estimate the resulting double-hurdle treatment model, which accounts for unobserved heterogeneity in the retention effect. A nearest-neighbour matching estimator is also implemented. Same-grade comparisons.	The paper estimated the causal effect of repeating kindergarten on the retained children's academic performance. Repeating kindergarten has positive effects on the retained children's later academic performance (i.e. the retained children would do worse in terms of the first- and third-grade test scores, were they socially promoted). Results also suggest that these effects diminish over time. Comparison of the results from the control function and matching approaches shows that unobserved child, family, and school characteristics that affect a child's probability of repeating kindergarten also affect his academic performance. Kindergarten school.
Eide, E. and Showalter, M. (2001) in *Economics of Education Review*	US. High School and Beyond (HSB) dataset. 1980 (sophomore high school), and follow ups 1982, 1984, 1986 and 1992. $N = 7809$	OLS by gender and race (white vs. black) and IV estimation, with the exogenous variation across states in kindergarten entry dates as the instrument. Same-age comparisons.	For all demographic groups, the OLS estimates showed a statistically significant positive correlation between retention and dropping out of high school and a statistically significant negative correlation between retention and post-high school labour market earnings. However, the IV estimates indicated that, for whites, grade retention might have some benefit to students by both lowering dropout rates and raising labour market earnings, although none of the coefficients were statistically significant. Primary school.

(*Continued*)

Table A1. *Continued.*

Longitudinal databases			
Author (year)	**Country and database**	**Methodology**	**Main results**
Eren *et al*. (2017) in *Journal of Public Economics*	US, Louisiana. Administrative records of the Louisiana Department of Education (LDOE) from 1999 through 2012. $N = 155{,}000$	Regression discontinuity design Same-grade comparisons.	The results indicate that potential grade retention, even at fourth grade, increases the propensity that a student drops out of school at a later point in time. In addition, eighth grade remedial education assignment in the form of summer school appears to provide a positive benefit by decreasing the likelihood that a student later drops out. Primary and secondary school.
Ferreira *et al*. (2018) in *IZA Discussion Paper*	Colombia. The first is a dataset from the Colombian Inspectorate of Education about the centralized exam conducted among 2.7 million pupils in their last year of secondary education (11th grade).	Difference-in-differences Same-grade comparisons.	Retained students improve their performance on language but not on math test scores. Secondary school
Figlio, D. and Özek, U. (2020) in *Journal of Public Economics*	United States of America. 12 Florida county-level school districts. Student-level administrative data that cover the school years between 2000−2001 and 2011−12. $N = 392{,}121$	Regression discontinuity design (RDD). Same-grade comparisons.	Results for English learners show that retention in the third grade, together with instructional support, improves the English skills of these students to a great extent. It also decreases the time by half of proficiency and it also reduces the probability of taking remedial English courses afterwards. Besides, it increases the likelihood of taking advanced courses in math and science as well as taking college credit-bearing courses in high school for English learners. The authors do not find any negative effects on disciplinary matters or absences for these English students. Primary and secondary school.

(*Continued*)

Table A1. *Continued.*

Longitudinal databases

Author (year)	Country and database	Methodology	Main results
García-Pérez *et al*. (2014) in *Applied Economics*	Spain. PISA 2009 $N = 25{,}887$	Switching regression model (SRM), with and without instrument by maximum likelihood Same-age comparisons.	The main objective of the study is to estimate the effect of grade retention on educational attainment. Grade retention decreases test score in math of repeaters in 54 points. Moreover, the results show that if a student repeated at the primary level, she will suffer a causal decrease in her performance, but this situation could be even worse if this student was subjected to a second-grade retention in secondary school. Primary school.
Gary-Bobo *et al*. (2016) in *Quantitative Economics*	France. 1995 secondary education panel of the French Ministry of Education (DEPP Panel 1995). Grade 6 to grade 9. 1995–2001. $N = 17{,}830$	Preliminary IV and a multi-stage model of human-capital accumulation with a finite number of types representing unobserved individual characteristics. Same-grade comparisons.	The paper studied the treatment effect of grade retention taking unobserved heterogeneity and the endogeneity of grade repetitions into account. Estimation results showed that the average treatment effect on the treated (ATT) of grade retention on test scores is small but positive at the end of grade 9. The ATT of grade retention is higher for the weakest students. We also show that class size is endogenous and tends to increase with unobserved student ability. The Average Treatment Effect (ATE) of grade retention is negative, again with the exception of the weakest group of students. Grade repetitions reduce the probability of access to grade 9 of all student types. Secondary school.

(*Continued*)

Table A1. *Continued.*

Longitudinal databases			
Author (year)	**Country and database**	**Methodology**	**Main results**
Glick, P. and Sahn, D. (2010) in *World Bank Economic Review*	Senegal. Program on the Analysis of Education Systems of the Conference of Francophone Ministers of Education (PASEC). Test scores in second grade and follow up. Senegal Household Education and Welfare (EBMS) survey.	Regression analysis with multiple test observations using IV to correct for measurement error. School fixed effects and school random effects to correct for heterogeneity across schools. Same-grade comparisons.	The author found that, conditional on academic ability, repeating a grade has a negative impact on school progression, implying that the private costs associated with stringent repetition policies exacerbate the negative effects on attainment of poor early academic outcomes. Primary school.
Goos *et al*. (2013b) in *Journal of School Psychology*	Flemish educational system.	Three-level curvilinear growth curve models with a PSM as a first step. Same-age comparisons.	Results showed that first-grade retention was less helpful for struggling students than generally thought by parents and educators. First-grade repeaters seemed to outperform their equally at-risk but continuously promoted grade-mates in math and reading fluency during the retention year, but that this benefit seemed very short-lived. In fact, it even seemed that this benefit had already disappeared completely in second grade. Primary school.
Greene, J. and Winters, M. (2007) in *Education Finance and Policy*	US. Florida Department of Education. 2001–2002 to 2004−2005. grade. $N = 73{,}695$; $N = 7087$	IV (test-based promotion policy) for across year comparisons (students who were essentially separated by the year in which they happened to have been born) and RDD. Same-grade comparisons.	The authors evaluated Florida's test-based promotion policy one and two years after its initial implementation. Results showed that retained students slightly outperformed socially promoted students in reading in the first year after retention, and these gains increased substantially in the second year. Result interpreted as average treatment effect across all of the interventions under the policy (i.e. summer school). Primary school.

(*Continued*)

Table A1. *Continued.*

Longitudinal databases			
Author (year)	**Country and database**	**Methodology**	**Main results**
Hill, A. (2014) in Economics of Education Review	US. National Longitudinal Study of Adolescent Health (Add Health). Grades 7–12 during the 1994−1995. $N = 6341$.	Regression longitudinal analysis using individual and school-specific course fixed effects to control for ability and course selection. Same-grade comparisons.	This paper investigated the causal effect of course repeaters on other students taking the course for the first time. An increase in the share of repeaters in a high school mathematics course leads to a significant increase in the probability of course failure for first-time course-takers. Results also show that course repetition externalities may be distinct from low- ability peer effects. Secondary school.
Hughes, J. *et al*. (2010) in *Educational Evaluation and Policy Analysis*	US. School longitudinal data. Texas Assessment of Knowledge and Skills: Reading and Math (TAKS). 2003/2004. $N = 769$	Authors aim to correct selection bias using propensity score matching (PSM) and generalized estimating equations (GEE). Same-grade comparisons.	Results showed that students who are retained in first grade are more likely to pass these tests (third grade) than they would have been if they had been promoted to second grade. The positive association between retention and math scores was significant while the association was marginally significant for reading scores. Primary school.

(*Continued*)

Table A1. *Continued.*

Longitudinal databases			
Author (year)	**Country and database**	**Methodology**	**Main results**
Im, M. *et al*. (2013) in *Journal of School Psychology*	US. Longitudinal school data in three school districts in Texas. 2000. Grades 1 to 5. Reading and Math tests. $N = 784$	Authors aim to correct selection bias using propensity score matching (PSM) and piecewise quadratic latent growth models. Same-grade comparisons.	Results showed that students who are retained in grades 1 to 5 are performing in middle schools as well as their propensity matched, continuously promoted peers, both academically and in terms of behavioural engagement and student-reported school belonging. Retention did not appear to offer any advantage to these students, nor did it impede their performance in middle school. Accepting the assumption that the close propensity matching mimics the results of a randomized, experimental trial, if retained students had been promoted in grades 1 to 5, they would be performing just as well but would be one year closer to high school completion. Primary school.
Jacob, B. (2005) in *Journal of Public Economics*	US-Chicago. Panel of student-level administrative data. CPS 1993–2000. $N = 400{,}000$	Difference-in-difference estimator Same-grade comparisons.	Effect of incentives on academic performance. Impact of high-stakes testing in Chicago Public Schools. The author used the argument of deterrence effects of retention: one might believe that the prospect of sanctions for low performance may lead to higher achievement by increasing student effort, raising parent participation or improving curriculum and pedagogy. The results of this analysis suggest that the high-stakes testing policy led to substantial increases in math and reading performance on the high-stakes test. Primary school.

(*Continued*)

Table A1. *Continued.*

Longitudinal databases

Author (year)	Country and database	Methodology	Main results
Jacob, B. and Lefgren, L. (2004) in The Review of Economics and Statistics	US. Administrative data from Chicago Public Schools (CPS). 3rd and 6th grade students from 1993−1994 to 1998−1999. N = 147,894	RDD with IV estimates (Probability of being retained in 3rd grade and 6th grade as the instruments). Same-grade comparisons.	Authors found that retention has no negative consequences on the academic achievement of students retained in 3rd grade, actually increasing performance in the short run. Retention increases achievement for third-grade students and has little effect on math achievement for sixth-grade students. They also presented results on summer school effects. Primary School.
Jacob, B. and Lefgren, L. (2009) in American Economic Journal: Applied Economics	US. CPS. 1997, 1998 and 1999. N = 20,000	Authors use plausibly exogenous variation in retention generated by a test-based promotion policy to assess the causal impact of grade retention on high school completion. RDD and IV estimates. Same-grade comparisons.	The authors reported that retention among sixth-grade students does not affect the likelihood of high school completion, but that retaining low-achieving eighth-grade students in elementary school substantially increases the probability that these students will drop out of high school. Primary School.
Kretschmann, J. *et al*. (2019) in *Journal of Educational Psychology*	German students over three years of secondary school sixth grade, $N = 3288$	Authors use propensity score matching on baseline measures of the dependent variables, and covariates associated with the risk of retention: cognitive ability, academic performance, and family background. Same-age comparisons.	Authors found no positive effects of retention on students' academic self-concept, a decrease in their academic self-concept, interests and learning motivation during the last months spent in the original class, just before retention. These negative effects endured one year after the episode of retention but decreased two years after grade retention. Primary school.

(*Continued*)

Table A1. *Continued.*

Longitudinal databases			
Author (year)	**Country and database**	**Methodology**	**Main results**
Lamote, C. *et al*. (2014) in *Educational Studies*	Flanders. Flemish longitudinal data (LOSO-project). N = 3,900	Authors use PSM so to compare treatment and control groups (same-grade comparison). Then, they use growth curve analysis. Same-grade comparisons.	Authors focused on the effect of grade retention in Grade 8 on language achievement and academic self-concept. Results show that grade retention had no negative effect on achievement in the short term (year of retention) and is positive in self-concept. On the other hand, in the long run it does have strong negative effects on achievement and no effect on self-concept. Secondary school.
Lorence, J. (2014) in *Social Science Research*	US. Texas Education Agency (TEA). Annual individual-level data of all students enrolled in Texas public schools from 1994 through 2002. $N = 38{,}000$	Logistic regression, multiple regression; propensity score matching (PSM) and two-level hierarchical linear model. Same-grade comparisons.	The study used propensity score matching to assess the causal effect of third grade retention on reading performance in later grades. Same- grade comparisons show that third graders failing the state-mandated reading test who repeated the grade consistently outperformed in later grades the socially promoted children who also failed the third-grade test. The results are consistent with findings from other recent studies that suggest that grade retention in third grade may help increase student achievement. Primary school.
Mahjoub, M. (2017) in *Education Economics*	France. Direction de l'Evaluation, de la Prospective et de la Performance (DEPP): the statistics department within the French Ministry of Education DEPP Panel 95 data. $N =$ 12,000	IVs and matching estimators. Same-grade comparisons.	The two methods give a positive effect of grade repetitions, between 10% and 25% of the test-score gain's standard deviation. A grade repetition improves the probability to graduate from junior high school by 2.5 probability points. Secondary school.

(*Continued*)

Table A1. *Continued.*

Longitudinal databases			
Author (year)	**Country and database**	**Methodology**	**Main results**
Manacorda, M. (2012) in R*eview of Economics and Statistics*	Uruguay. Administrative longitudinal microdata on junior high school students. Grades 7 to 9. 1996–2001. $N =$ 99,729	RDD based on a promotion rule (more than three failed subjects implies grade failure) and IV estimator. Same-grade comparisons.	Results showed that grade failure leads to dropout and lower educational attainment four to five years after failure on the order of −0.8 school years. When accounting for the potential non- random selection of students around the discontinuity threshold using reasonable (although untestable) assumptions, the author finds estimates for the effect of grade failure that are negative and in the order of −0.2 school years. Secondary school.
Moser *et al*. (2012) in *Journal of Educational Psychology*	US. From a large multi-ethnic sample of children who were below the median in literacy at school entrance, 363 children who were either promoted ($n = 251$) or retained ($n = 112$) in first grade $N = 784$	Using longitudinal growth curve analysis with PSM to create comparable groups. Same-grade comparisons.	For both math and reading achievement scores, there is an initial advantage in achievement for students' repeated first grade scores compared to their promoted peers' first grade scores. However, this effect dissipates over time. Primary school.
Namen, O. (2018) in Unpublished manuscript, Department of Economics, Universidad del Rosario and Innovations for Poverty Action (IPA).	Chile. The first dataset corresponds to administrative student records collected by the Ministry of Education of Chile and the second dataset corresponds to standardized test scores from the national exam SIMCE taken yearly by all students in fourth grade. $N = 731{,}336$	Difference-in-differences Same-grade comparisons.	The study found that retained students improve their performance on language but not on math test scores. Primary school.

(*Continued*)

Table A1. *Continued.*

Longitudinal databases			
Author (year)	**Country and database**	**Methodology**	**Main results**
Nunes *et al*. (2018) in *Applied Economics*	Portugal. Administrative database managed by the Portuguese Ministry of Education containing information about students in public schools.	Propensity score matching. We estimate both the average treatment effect (ATE) and the average treatment effect on the treated (ATET). Same-grade comparisons.	The results suggest that in some situations retentions may have on average a positive impact on future achievement. However, in the cases where statistically significant impacts are found, the estimated magnitudes are relatively small. Our results are relevant for countries with high retention rates that are considering alternative educational policies to promote students' achievement. The impact of retention at the 4th grade on the scores obtained on the 6^{th} grade exams, controlling for the level of ability at the moment of retention. Primary school.
Ou, S. and Reynolds, A. (2010) in *Educational Evaluation and Policy Analysis*	US. Chicago Longitudinal Study (CLS)−2005. $N = 1367$	Probit regression analysis and propensity score matching so to correct for selection bias. Same-age comparisons.	Results showed that grade retention is significantly associated with lower rates of participation in postsecondary education above and beyond the effects of family demographics and early school achievement. Primary and secondary schools.
Pereira, M. and Reis, H. (2014) in Economic Bulletin and Financial Stability Report Articles, Banco de Portugal.	Portugal. PISA 2003 & 2009	Heckman control function method + IV Same-age comparisons.	Academic performance (reading and mathematics) at a later stage of basic education is negatively affected by repeating at an early stage. the short-term effects of repeating at a later stage are positive, although small. Primary school.

(*Continued*)

Table A1. *Continued.*

Longitudinal databases			
Author (year)	**Country and database**	**Methodology**	**Main results**
Roderick, M. and Nagaoka, J. (2005) in *Educational Evaluation and Policy Analysis*	US. Administrative data from CPS. Third- and sixth-grade students from 1997–2000.	RDD with IV estimates to address selection effects. Same-grade comparisons.	The study found a small positive impact on the performance of retained third- and sixth-grade students relative to promoted students in the year that the students were retained, but within the next two years, these gains disappeared among the third graders and were reversed among the sixth graders, such that the retained sixth graders had actually fallen behind their promoted counterparts. Primary school.
Winter, M. and Green, P. (2012) in *Education Finance and Policy*	US. Florida Department of Education's PK-20 Education Data Warehouse. $N = 75{,}000$ (pooled sample)	Two-stage least squares model. regression discontinuity implemented via IVs estimation. Same-age and same-grade comparisons.	Authors find evidence of substantial short-term gains in both math and reading achievement. These positive effects fade out over time and become statistically insignificant within five years when retained students are compared to peers of the same age, but remain substantial when retained students are compared to peers in the same grade. They also find that third grade retention and remediation substantially reduce the probability of being retained in later grades and has no clear impact on the probability of graduating from high school. Primary school.
	US. Florida Department of Education. 2002–03 to 2008–2009. Third grade to seventh grade.	RDD. Same-grade comparisons.	Authors analysed the causal effect of remediation policy (retention, high quality teacher and summer school) 5 years after intervention. Exposure to these interventions has a substantial positive effect on student achievement in math, reading, and science in the years immediately following the treatment and dissipates over time. Authors point out that results apply only to this type of program and similar ones. Primary school.

(*Continued*)

Table A1. *Continued.*

Longitudinal databases

Author (year)	Country and database	Methodology	Main results
Xinxin, C. *et al.* (2010) in *International Journal of Educational Development*	China. The data used in this paper come from a survey executed by the authors in 2006. The survey was designed specifically to examine the changes in school achievement of children before and after they repeated at least one grade. $N = 1649$ in 36 elementary schools.	Differences-in-differences, propensity score matching and differences-in-differences matching approaches. Same-grade comparisons.	Results from the multivariate analysis consistently show that there is no significant positive effect of grade retention on school performance of the students. In fact, in some cases (e.g. for the students who repeat grade 2), grade retention is shown to hurt school performance. Primary school.

Cross-sectional

Author (year)	Country and database	Methodology	Main results
Belot, M. and Vandenberghe, V. (2014) in *Education Economics*	French speaking community of Belgium. PISA 2003, 2006. $N = 3700$ (control group N = 40,000).	Natural experiment: reintroduction of grade retention in 2001. DiD comparing changes in Belgium pre and post reform with countries in the control group. Same-grade comparisons.	Authors evaluated the possible threat (or motivational) effects of a grade retention policy. They showed that the typical grade attained at age 15 has *decreased* with the re-introduction of grade retention sanctions at the end of grade 7. They fail to find any statistically significant improvement of grade 10 test scores. There is no evidence supporting the existence of 'threat' benefits of grade repetition. Secondary school.

Table A2. Summary of Grade Retention Effects Considering the Outcome of Interest as Well as the Educational Level Where the Policy Is Applied.

Author(year)	Level(of grade retention)	Impact of grade retention on: (*)				
		Test scores		Graduation rates	School dropout	Other educational outcomes	Employment outcomes
		Short-Term	Long-Term				
Dong (2010)	Kindergarten	++					
Fruehwirth *et al*. (2016)	Kindergarten	–	-				
Alet *et al*. (2013)	Primary school	++	-				
Andrew (2014)	Primary school			–			
Babcock and Bedard (2011)	Primary school						+
Bhattacharya (2007)	Primary school	-					
Cabrera-Hernandez (2016)	Primary school				+		
D'Haultfoeuille (2010)	Primary school	+					
Diris (2017)	Primary school	–					
Eide and Showalter (2001)	Primary school				n.s		+
García-Pérez *et al*. (2014)	Primary school	-	–				
Glick and Sahn (2010)	Primary school				+		
Goos *et al*. (2013b)	Primary school	+				-	
Greene and Winters (2007)	Primary school	+					
Hughes *et al*. (2010)	Primary school	+					
Im *et al*. (2013)	Primary school	n.s					
Jacob (2005)	Primary school	++					
Jacob and Lefgren (2004)	Primary school	++					
Jacob and Lefgren (2009)	Primary school			n.s	++		
Kretschmann *et al*. (2019)	Primary school					–	
Moser *et al*. (2012)	Primary school	+					
Namen (2018)	Primary school	+					
Lorence (2014)	Primary school	+					
Nunes *et al*. (2018)	Primary school	+					
Pereira and Reis (2014)	Primary school	+	-				

(*Continued*)

Table A2. *Continued.*

Author(year)	Level(of grade retention)	Impact of grade retention on: (*)					
		Test scores		Graduation rates	School dropout	Other educational outcomes	Employment outcomes
Roderick and Nagaoka (2005)	Primary school	-		+			
Schwerdt *et al*. (2017)	Primary school	++	n.s				
Winters and Green (2012)	Primary school	++	n.s				
Xinxin *et al*. (2010)	Primary school	-					
Eren *et al*. (2017)	Primary and secondary school				+		
Figlio and Özek (2020)	Primary and secondary school	++	++			++	
Ou and Reynolds (2010)	Primary and secondary school					–	
Belot and Vandenberghe (2014)	Secondary school		n.s				
Brodaty *et al*. (2014)	Secondary school			–			
Cockx *et al*. (2019)	Secondary school			-	+		
Ferreira *et al*. (2018)	Secondary school	++					
Gary-Bobo *et al*. (2016)	Secondary school	+		-			
Hill (2014)	Secondary school				+		
Lamote *et al*. (2014)	Secondary school	n.s	–				
Mahjoub (2017)	Secondary school	++					
Manacorda (2012)	Secondary school		-		+		
Brodaty *et al*. (2013)	Secondary school and higher education						-

* The signs refer to magnitude of the corresponding significant effect:
'++' if strong positive effect,
'+' if positive effect,
'-' if negative effect,
'—' if strong negative effect.
n.s indicates that the effect is non-significant.