Education Economics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/cede20

Evaluating the 'threat' effects of grade repetition: exploiting the 2001 reform by the French-Speaking Community of Belgium

Michèle Belot [a] & Vincent Vandenberghe [b]

[a] Nuffield Centre for Experimental Social Sciences (CESS), Nuffield College, Oxford University, Oxford, UK

[b] Institut de Recherches Economiques et Sociales (IRES), Université catholique de Louvain (UCL), Place Montesquieu 3, B-1348, Louvain-la-Neuve, Belgium

Available online: 06 Sep 2011

To cite this article: Michèle Belot & Vincent Vandenberghe (2011): Evaluating the 'threat' effects of grade repetition: exploiting the 2001 reform by the French-Speaking Community of Belgium, Education Economics, DOI:10.1080/09645292.2011.607266

To link to this article: http://dx.doi.org/10.1080/09645292.2011.607266

Education Economics
2011, 1–17, iFirst Article

Evaluating the 'threat' effects of grade repetition: exploiting the 2001 reform by the French-Speaking Community of Belgium

Michèle Belot[a] and Vincent Vandenberghe[b]*

[a]*Nuffield Centre for Experimental Social Sciences (CESS), Nuffield College, Oxford University, Oxford, UK;* [b]*Institut de Recherches Economiques et Sociales (IRES), Université catholique de Louvain (UCL), Place Montesquieu 3, B-1348, Louvain-la-Neuve, Belgium*

(*Received 3 August 2010; final version received 9 August 2011*)

Like active labour market programmes, grade repetition could generate two types of effects: better/worse outcomes due to programme participation (i.e. the fact that pupils repeat a particular grade). This is what the existing literature on grade repetition has focused on. Another potential outcome is the 'threat' effect of grade repetition. Pupils and/or their family could make significant efforts to avoid grade repetition and its important opportunity cost. Learning effort by pupils could be a function of the risk of grade repetition. This paper attempts to assess that relationship by exploiting a reform introduced in 2001 in the French-Speaking Community of Belgium, synonymous with a reinforced overall threat of grade repetition. The possibility to impose grade repetition sanctions and the end of grades 8–12 has always existed, but in year 2001, policy makers reinstated the possibility to repeat grade 7, putting an end to the regime of 'social promotion' applicable to that grade since 1995. We use data from two waves of the Programme for International Student Assessment study (corresponding to periods before and after the reform) to evaluate the medium-term effects of this reform. The first measure of performance we consider is the position in the curriculum (or grade) reached at the age of 15, and we show that it deteriorated after 2001. We also consider the reform's impact on test scores. Focusing on grade 10, we fail to verify the necessary condition for grade repetition threat to lead to higher test scores. The tentative conclusion is that an enhanced threat of grade retention after 2001 did not lead to better medium-term outcomes, even among the segments of the population the most at risk of grade repetition.

JEL classifications: I20; I28; H52

Keywords: grade retention; educational attainment; threat effects

1. Introduction

Grade repetition (or retention) is a contentious issue. Some countries privilege a system of 'social promotion', which allows pupils to be promoted to higher grades independently of their performance, while other countries have instituted more or less strict policies of grade retention, conditioning promotion to higher grades on educational achievements. As a consequence, there is a considerable variation in grade retention

*Corresponding author. Email: vincent.vandenberghe@uclouvain.be

ISSN 0964-5292 print/ISSN 1469-5782 online

DOI: 10.1080/09645292.2011.607266
http://www.informaworld.com

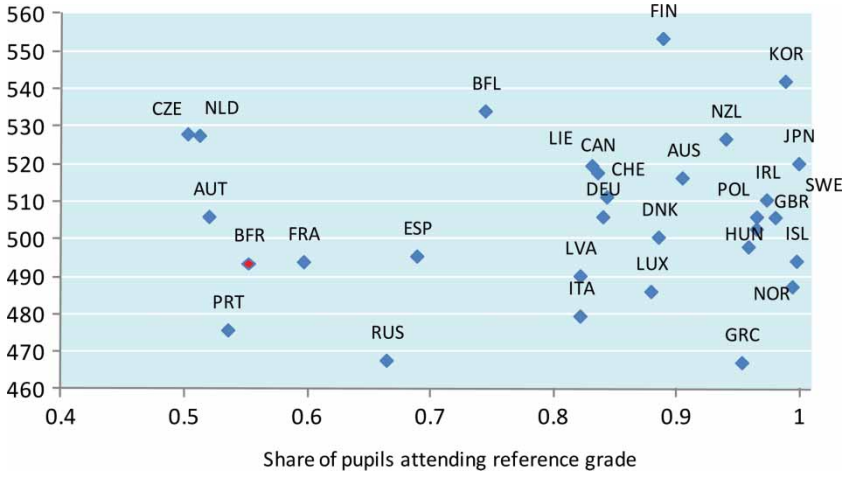

Figure 1. Average score in math and share of pupils aged 15 attending reference grade (Grade 10 in most countries and grade 9 otherwise; the grade of reference is identified as the most attended grade among 15-year-olds who participated to PISA). Year 2006.
Note: ARG, Argentina; AUS, Australia; AUT, Austria; AZE, Azerbaijan; BFR, French-Speaking Community of Belgium; BFL, Flemish-Speaking Community of Belgium; BGR, Bulgaria; BRA, Brazil; CAN, Canada; CHE, Switzerland; CHL, Chile; COL, Colombia; CZE, Czech Republic; DEU, Germany; DNK, Denmark; ESP, Spain; EST, Estonia; FIN, Finland; FRA, France; GBR, United Kingdom; GRC, Greece; HKG, Hong Kong-China; HRV, Croatia; HUN, Hungary; IDN, Indonesia; IRL, Ireland; ISL, Iceland; ISR, Israel; ITA, Italy; JOR, Jordan; JPN, Japan; KGZ, Kyrgyzstan; KOR, Korea; LIE, Liechtenstein; LTU, Lithuania; LUX, Luxembourg; LVA, Latvia; MAC, Macao-China; MEX, Mexico; MNE, Montenegro; NLD, The Netherlands; NOR, Norway; NZL, New Zealand; POL, Poland; PRT, Portugal; QAT, Qatar; ROU, Romania; RUS, Russian Federation; SRB, Serbia; SVK, Slovak Republic; SVN, Slovenia; SWE, Sweden; TAP, Chinese Taipei; THA, Thailand; TUN, Tunisia; TUR, Turkey; URY, Uruguay; USA, United States of America.
Source: PISA, OECD.

rates across Organisation for Economic Cooperation and Development (OECD) countries (Figure 1). Countries/entities such as the Netherlands (NLD), Austria (AUT), Portugal (PRT) and the French-Speaking Community of Belgium (BFR) have relatively high rates of grade retention (going up to 50% of pupils having repeated a year or more by the time they reach the end of compulsory schooling); while countries such as Norway (NOR), Sweden (SWE), Japan (JPN) and Great Britain (GBR) have no grade retention at all.

Analysts disagree as to which policy is the most effective. The point we raise in this paper is that this debate about the pros and cons of grade repetition would gain some insights by considering recent developments of the evaluation literature on active labour market policies (ALMPs). These are government programmes aimed at helping the unemployed find work. Many of them consist of imposing that benefit recipients with poor (re)employment prospects participate to training schemes such as classes and apprenticeships. ALMPs are motivated by a need to upgrade the skills of unemployed individuals in order to increase their employability.

We argue here that grade repetition bears a strong likeness to ALMPs. First, grade repetition is motivated by the need to upgrade the (basic) skills of less-able individuals (i.e. pupils) in order to increase their future prospects (i.e. their chance to grasp the

more advanced curriculum taught in higher grades). The advocates of grade repetition insist on giving weak students another opportunity to acquire the necessary skills before being passed on to a level for which they would not otherwise be prepared. Second – and more importantly in the context of this paper – the evaluation literature on ALMPs has established a distinction between (i) the treatment or *ex post* effects of programme participation and (ii) the 'threat' or *ex ante* motivational effects of programmes. Whereas the existing literature on grade repetition has largely covered the first aspect, it has, so far, largely neglected the threat effects of grade repetition. This paper intends to fill that void. It evaluates these threat effects by exploiting a reform introduced in 2001 in the French-Speaking Community of Belgium. That reform resulted into an enhanced overall threat of grade repetition. Although the possibility to impose grade repetition in grades 8–12 always existed, the year 2001 brought a significant regime change as policy makers reinstated the possibility to repeat grade 7 (first grade of secondary education), putting an end to 'social promotion' applicable to that grade since 1995.

The main results of the paper are that an enhanced threat of grade retention after 2001 did not lead to better outcomes, even among the pupils the most at risk of grade repetition (the 'borderline' students hereafter). The first measure of educational performance we consider is the position in the curriculum (or grade). We show that the typical grade attained at age 15 has *decreased* with the re-introduction of grade retention sanctions at the end of grade 7. The overall proportion of pupils who made it to grade 10 at the age of 15 (i.e. those with a no-grade-repetition record) fell by about 4 percentage points. The corresponding fall among borderline pupils (e.g. those with low-educated mothers or from low socio-economic background *a priori* more exposed to the risk of grade repetition) is estimated to be in the range of 10–14 percentage points. These results may seem trivial. However, we think that they are probably not a mechanical implication of the reform because the threat/incentive argument predicts an improvement of performance. The reform should have increased the proportion of pupils making it to grade 10 without grade repetition. The most fervent proponents of grade repetition think, indeed, that grade repetition sanctions bear some likeness to Cold-War arsenal: their sole existence suffices to significantly alter human behaviour. This paper contains evidence that they may be overoptimistic in that respect.

We consider a second, *a priori* more natural, measure of outcome: test scores. These are reflecting the actual cognitive skills attained by pupils at a certain grade. We focus on those of the pupils with no grade repetition record, thus attending grade 10 at the age of 15. We have stated above that the 2001 reform reduced the proportion of those pupils, presumably by sorting out the less-able ones. The change in grade 10 average score is thus likely to reflect a (presumably positive) *screen out effect*. But, it should also reflect the *threat effects* we are interested in, singularly among the borderline students forming the lower end of the grade 10 test score distribution. Both effects should *a priori* reinforce each other and lead to an improvement in the average grade 10 test scores. But, our results point at the absence of statistically significant improvement. We then conclude that there has been no medium-to-long-term benefit to enhanced grade repetition threat.

The remainder of this paper is organised as follows. Section 2 develops the comparison between grade repetition and ALMPs and the way these policies have been evaluated in the literature. Section 3 presents the 2001 reform in the French-Speaking Community of Belgium and documents its interest as a source of exogenous variation of the threat of grade repetition. Section 4 presents the results of the empirical analysis of threat effects. Section 5 concludes.

2. Participation *vs*. threat effect

'*Is the threat of reemployment services more effective than the services themselves*'? This is the question asked by Black et al. (2003) in their seminal paper. It epitomises the distinction now commonly made by labour economists and employment-policy evaluators between the benefits people derive from programme participation (e.g. a training programme for the long-term unemployed) and the way they respond *ex ante* to the cost or discomfort associated to mandatory participation: something known in that literature as the 'threat' effect of the programme.

Threat effects are probably driven by the opportunity costs associated to programme participation (less leisure, locking-in problems, etc.). These seem to suffice to entice unemployed individuals to raise their search effort prior to participation and rapidly move out of unemployment (Rosholm and Svarer 2004, 2008; Cockx and Dejemeppe 2007). We argue here that pupils (and/or their family) could also raise their efforts and study harder to avoid the opportunity cost of grade repetition that is *a priori* high. A grade repetition sanction means indeed that a whole extra year is required to preserve the possibility to obtain a certain diploma.

There is now ample evidence that the participation to ALMPs has small, and in some cases even adverse, effects in terms of increasing job-finding rates for the long-term unemployed; see, for example, Heckman, Lalonde, and Smith (1999). By contrast, the ALPMs evaluation literature produces robust evidence of positive threat effects. Black et al. (2003) conclude that threat is the main gain of a training programme imposed to American unemployed with the lowest reemployment prospects. Other researchers have shown that the prospect of mandatory participation to ALMPs (sometimes just receiving a letter of notification) impacts the unemployment exit rate as much as programme participation (Rosholm and Svarer 2004; Geerdsen 2006; Geerdsen and Holm 2007).

So far, the education literature has exclusively focused on the participation (or *ex post*) effect of grade retention, namely its consequences on final attainment. Holmes (1989), in a large meta-analysis, finds that, on average, later test scores of children retained are 0.19–0.31 standard deviations lower than those of similar children progressing normally through school. The same negative results are reported in a subsequent meta-analysis by Jimerson (2001). Belgian evidence on this is surprisingly limited. The only published study we came across is the one by Goos et al. (2010). This said, it also finds that first-grade repeaters would have performed better and would have shown a similar or even better outcome had they been promoted to second grade.

There is also a large amount of evidence of a negative relationship between retention (i.e. participation) and high school dropout (e.g. Grissom and Shepard 1989; Roderick 1994; Jimerson 1999).

Part of that literature tries to address the endogeneity of grade retention by providing quasi-experimental evidence of the effects of grade retention. Eide and Showalter (2001) use the variation in the age of entry into kindergarten across US states as an instrument for grade retention. They find that for white students, grade retention may have some benefits by both lowering dropout rates and raising labour market earnings, although their instrumental variable (IV) estimates tend to be statistically indistinguishable from zero. Three studies (Jacob and Lefgren 2004, 2009; Roderick and Nagaoka 2005) exploit a discontinuity in the retention decision under Chicago's high-stakes testing policy introduced in 1996–1997. The policy established a much stronger relationship between low scores in a single standardised test and the probability of

grade retention. Using a regression discontinuity design, these studies evaluate the *ex post* effects of grade retention on pupil performance at different points in time. Jacob and Lefgren (2004) find no systematic differences in performance between retained and promoted students in the short run. Roderick and Nagaoka (2005) show that third-grade students who were retained do not yield higher language test scores 2 years after the retention and that retained sixth graders had lower achievement growth. Finally, Jacob and Lefgren (2009) find that grade retention leads to a modest increase in the probability of dropping out for older students, but has no significant effect on younger students. Manacorda (forthcoming), exploiting the discontinuity induced by a rule establishing automatic grade repetition for Uruguayan pupils with more than three failed subjects, shows that grade repetition leads to substantial dropout and lower educational attainment even 4–5 years after repetition first occurred.

In short, these studies nicely evaluate the causal effects of being retained. Like many ALMPs, evaluation studies seem to produce mixed evidence about the *ex post* benefits of grade retention (those stemming from participation). However, these studies do not evaluate the possible threat (or motivational) effects of a grade retention policy. The work by Jacob (2005) is a noteworthy exception. It is the most closely related paper to ours, as it almost exclusively focuses on what its author calls the 'incentive' effect of retention.[1] Using before-and-after mathematics and reading test scores of pupils affected by the above-mentioned third, sixth and eighth grade Chicago high-stake testing policy,[2] he documents some evidence of motivational effects of such a policy, mainly for eighth graders. He finds little evidence of benefits that can be ascribed to student-oriented incentives across grades 3–7.

Although it would be hard from a policy-making perspective to justify a policy of grade retention solely because of positive threat effects, rather than because it is directly beneficial to those who are retained, we believe that it is worthwhile investigating whether there is any evidence of positive threat/motivational effects in the first place.

3. Exploiting the French-Speaking Community reform to assess the threat/motivational effects of grade retention

Grade retention has existed for a long time in Belgian secondary schools and has been extensively used to sanction students with weak results at the end of all grades (7–12). It is particularly frequent in the French-Speaking Community[3] (Figure 1). The retention decision in French-speaking schools is based on the teachers' assessment of the pupil's ability of passing to a higher grade. There is no standardised test used across schools, nor is there a clearly defined threshold to determine whether a pupil should be retained or not. All pupils do take exams at the end of the school year for each subject, and the retention decision is made after these exams have been taken.

Opponents to grade retention succeeded in 1995 in almost eliminating grade repetition at the end of grade 7 (first year of secondary education). However, the possibility to impose grade repetition sanctions and the end of grades 8–12 remained unaltered. The argument for suppressing grade retention at the end of grade 7 at the time was that the entrance into secondary education involves a sharp change in the schooling environment – moving from one teacher to a range of teachers – and that pupils may need time to adjust to this new environment. From 1995 to 2001, the main rule was that no grade retention was allowed at the end of grade 7, a decision that translated into a sharp fall of the share of 'repeaters' (Figure 2). Pupils could only possibly repeat grade 7 upon agreement between parents and teachers. This is why on Figure 2 one

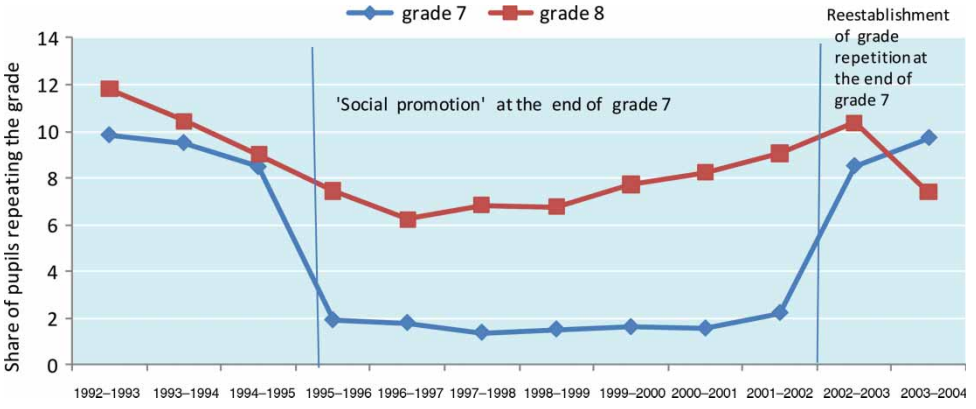

Figure 2. Incidence of grade retention at grades 7 and 8.
Note: School year 1992–1993 to 2003–2004.
Source: French-Speaking Community of Belgium, Ministry of Education.

observes a persistence of grade retention at the end of grade 7 during the 1995–2001 period.

The proponents of grade retention made a successful comeback 6 years later. In September 2001, the decision was made[4] to re-establish the possibility of retaining pupils at the end of grade 7. In a few words, the 2001 reform was such that after the school year 2001–2002, it became possible to repeat grade 7 or grade 8, although not both.[5] The arguments in favour of grade repetition were essentially twofold. First, it could be that it is better to retain pupils earlier rather than later. Second, the re-introduction of a threat of retention in grade 7 could provide incentives for weaker pupils to put in more effort, something that would make sense in a context where pupils value the costs of repeating in the near future heavier than the costs of repeating later. It is important to stress that the post-2001 regime kept the incentives to perform in grade 8 and beyond as high as before the reform; those who passed grade 7 were exposed to the threat of having to repeat subsequent grades in the same way that they were under the system with no retention in grade 7. The main difference is that the reform introduced an additional evaluation and retention decision, *a priori* reinforcing the overall threat of grade repetition.

In truth, the reform did not represent a switch in one go from one extreme (no retention at any stage of the curriculum) to the other extreme (retention at all grades of the curriculum). However, the 2001 reform concerned one of the most important moments of a pupil's career. The marginal return to study effort during grade 7 could be higher than that during, say, grade 10 or 11. We also think that the most relevant evaluation works are those considering reforms that introduce changes at the margin, as most of the time policy-making consists of timid and relatively marginal moves.

Administrative data (Figure 2) show that the share of pupils repeating grade 7 rose from the school year 2002–2003 onwards. Quite strikingly, the same data also show that the share of students repeating grade 7 *or* grade 8 is substantially higher in 2002–2003 and beyond, not only meaning that the 2001 reform actually increased the (short-term) risk of grade retention, but also suggesting that it failed to entice pupils to work harder to avoid that sanction and its large opportunity cost.

At this stage of the paper, the main message is that the 2001 reform enables us to evaluate the effect of an additional intermediary retention sanction. Before 2001, the decision to retain pupils was delayed until grade 8. After 2001, pupils could already be retained at the end of grade 7. Hereafter, we exploit data from the Programme for International Student Assessment (PISA) study: the OECD's international standardised assessment administered to 15-year-old students. This allows us to investigate more carefully the medium-to-long-term[6] (causal) threat effects of the 2001 reform on (i) grade attainment and (ii) standardised test scores.

Note already that PISA only contains *same-age* test scores (all respondents are aged 15), meaning that we are not able to evaluate the (*ex post*) participation effects of grade repetition. Only an evaluation of test scores at equal level of curriculum (same-grade test scores of repeaters *vs*. nonrepeaters) would be informative about whether repetition is beneficial or not. Since in PISA there are differences in positions in the curriculum, the comparison of test scores between retained and nonretained pupils is not directly meaningful. By contrast, PISA test scores are perfectly suitable to assess the threat effects of grade repetition.

4. Empirical analysis

4.1 *Data, identification strategy and control group*

In each country that participated to PISA, standardised tests in Maths, Science and Reading literacy were administered to representative samples of pupils (we also run our analysis using the average score obtained by a respondent for each of these three topics). We have information from the PISA study before the reform (2003) and after the reform (2006), measuring, at the age of 15, grade attainment and cognitive performance. In addition, the PISA assessment contains many questions about attended grade and programme student background or school characteristics.

The 2003 wave includes pupils who experienced the social promotion regime during their first year of secondary education (no risk of retention in grade 7), and the 2006 wave includes pupils who were affected by the reform (risk of grade repetition).[7] Since we are looking at 15-year-olds, the grade reference for these pupils is grade 10 (grade that they would have attained without retention).

The identification of a causal effect of enhanced grade retention threats faces challenges. First, due to data constraints, we primarily use grade 10 test scores to detect an increase of effort that may have occurred during grade 7. There is no doubt that we are using students who have only been exposed to (varying degrees of) the 'threat' of grade repetition. By construction, their presence in grade 10 at the age of 15 means that they never repeated a grade. Our empirical configuration is thus ideal to separate threat from treatment effects. One may, however, argue[8] that the ideal regression to assess the benefits of enhanced threat would be to use test scores at the end of grade 7, before the grade retention decision is stated. But, the acquisition of cognitive skills is very cumulative; unless incentives to study in grades 8–10 were altered after the 2001 reform, one would still expect medium-term (i.e. grade 10) outcomes to reflect additional efforts made quite earlier. Second, resorting to between-country variance (Figure 1) is insufficient to properly identify the causal threat effect on scores of the grade repetition. Cross-country difference in terms of grade repetition incidence could be correlated with unobserved socio-economic or policy differences that also affect PISA scores. Third, changes observed within a country after a grade repetition

regime change may be driven by unobserved confounding factors that are correlated with scores, like a better/worse economic environment (insufficiently or inadequately captured by the observables available in PISA). Thus, ideally, the identification of the effects of grade retention requires not only an exogenous change in the threat of grade repetition, but also the existence of a counterfactual to account for time-related changes. This is why we resort to a difference-in-difference (DD) analysis, comparing the changes observed in the French-Speaking Community of Belgium to the changes observed in a *control group*. This approach also ensures that any systematic differences in the difficulty of the tests between the different years are netted out.

Of course, the challenge is to find an adequate control group. There is no ideal counterfactual country. We have experimented with different combinations of countries[9] and chose countries that satisfy the following criteria[10]: (i) located in Europe and (ii) where grade retention does not exist. This group comprises Great Britain, Greece, Hungary, Ireland, Iceland, Norway, Poland and Sweden.

Why Europe? This is because the 'parallel-trend' assumption is a critical assumption in a DD framework. Non-European OECD countries could be affected by overall trends that are not similar to those affecting Europe. The USA, Australia and New Zealand or Canada, for instance, are known for being relatively disconnected from Europe in terms of business cycle.

Why countries with no grade repetition policy? DD requires that nothing else than the intensity of grade repetition threat changes between the two groups between 2003 and 2006. We argue that, by definition, the countries forming our control group were unlikely to experience simultaneous changes in their grade assignment regime. By contrast, those with a grade repetition policy could have experienced changes of their own in the incidence of grade retention. There is, indeed, some evidence that many European countries are gradually limiting the possibility to impose grade repetition sanctions and try to promote the use of alternative 'remedial' strategies.

But, other sources of asymmetrical changes could exist. Educational policy comprises many more dimensions than the grade repetition regime, and these may have evolved asymmetrically, impacting country-specific test score trajectories. There is, however, little evidence suggesting that this was the case, at least at the very aggregate level. The third column of Table 1, for instance, suggests that overall test scores[11] gaps remained within the range of 2–3 points (unchanged in essence). The next two columns of Table 1 also show that school resources (student/teacher, computer/student ratio or share of certified teachers) did not evolve asymmetrically. Socio-economic characteristics influencing PISA scores could also have diverged between 2003 and 2006. For instance, the shares of students with low-educated mothers may have evolved in a diverging manner. Yet, descriptive statistics of Table 1 on key aspects of pupils' socio-economic background do not support this. In any case, we are able to account for socio-economic changes by adding control variables in our DD model (see Section 4.3 for more details). Finally, the use of a synthetic control group (i.e. formed by the aggregation of the European countries listed above) *a priori* reinforces the plausibility nothing else that the intensity of grade repetition threat has changed across the entities. There is, indeed, a reasonable chance that many of the confounding changes that exist in a *one-to-one* framework (i.e. based on the comparison of the French-Speaking Community of Belgium with just one other country) simply cancel out in the *one-to-many* setting used here.

In short, our study presents four main advantages. First, it examines the *threat* effects of grade retention; something that has received too little attention so far in

Table 1. French-Speaking community of Belgium *vs*. European controls, 2003 and 2006.

Entity	Year	Nobs	Average score[a]	Student/ teacher	Computer per student	Certified teachers (%)	Highest occupational index of father or mother	Highest diploma of mother[b]	Highest diploma of father[b]
Control (i.e. weighted average of Great Britain, Greece, Hungary, Ireland, Iceland, Norway, Poland and Sweden)	2003	39,228	496.68	11.43	0.20	0.94	49.58	3.39	3.34
	2006	45,571	496.13	11.80	0.19	0.94	49.95	3.54	3.47
French-Speaking Belgium	2003	3737	494.72	10.14	0.09	0.86	50.56	3.72	3.70
	2006	3733	493.40	9.90	0.11	0.78	50.62	3.64	3.66

Source: PISA, OECD.

Note: Descriptive statistics before (2003) and after (2006) natural experiment. All observations forming the control entity are weighted by the (inverted) country-specific PISA sample size. This is a way to reduce the risk that big countries – which sampled more pupils to achieve national representativeness – dwarf smaller ones.

[a]Based on students' average test score in Math, Sciences and Reading.

[b]ISCED scale: level 0, pre-primary education; level 1, primary education or first stage of basic education; level 2, lower secondary or second stage of basic education; level 3, (upper) secondary education; level 4, post-secondary nontertiary education; level 5, first stage of tertiary education; level 6, second stage of tertiary education.

the literature, although the threat/incentive argument lies at the core of many discussions and policy debates surrounding grade repetition. Second, the re-introduction of grade retention in 2001 provides a *natural experiment* to evaluate the effects of grade retention threat. Of course, the Belgian reform only re-introduces grade retention at the end of grade 7 (it has always been possible to impose grade repetition sanctions from grade 8 and beyond), so it is not clear that the overall incentives to study have been greatly affected. On the other hand, grade repetition as a threat *a priori* affects more students than grade repetition as a treatment. A small doze of uncertainty in the relationship between effort, ability, exam scores and actual grade repetition sanctions (combined with some risk aversion among pupils and their family) implies that the benefits of threat should be observed among a relatively large and diversified (in terms of socio-economic background) group of pupils. Only those at the extreme ends of the ability distribution (those who are certain to fail or pass) should not respond to an enhanced threat of grade repetition. Third, since schooling is compulsory until the age of 18 in Belgium, there is no possible dropout yet. This facilitates the comparison of scores pre- and post-reform. Fourth, our PISA data provide a *control group* suitable for carrying out a DD analysis, namely the European countries that do not have a grade repetition policy.

Our first measure of educational performance is the position in the curriculum (or grade). We show that the typical grade attained at age 15 has *decreased* with the re-introduction of grade retention at the end of grade 7. A smaller proportion of pupils reached grade 10 when grade retention was re-introduced at all grades. The reduction is even more significant among borderline pupils.

Our second outcome measure is performance at grade 10 (PISA test scores). These reflect the cognitive skills attained by pupils: a more objective measure of attainment.[12] We can test for the necessary condition for an enhanced threat of grade retention to have generated benefits, which is why it should be associated with an improvement in the distribution of scores. There are two possible mechanisms which could play a role. (i) A *screen out effect*: grade retention effectively prevents weaker students from gaining access to grade 10 at the age of 15. (ii) A *threat/incentive effect*: the threat of having to repeat a year should stimulate the pupils 'at risk' to work harder. Both mechanisms should lead to an improvement in the average test scores conditional on grade 10 attendance.

4.2 *Making it to grade 10 without grade repetition*

The first outcome variable we investigate is the presence in grade 10 at the age of 15 or, said differently, whether more or less pupils have succeeded in reaching that grade without repeating a grade. *Ceteris paribus*, the possibility of a sanction already at the end of grade 7 could have triggered an ounce of supplementary effort among students; something that may have put them on a (durably) successful learning curve synonymous with no grade retention.

A first indication that this was probably not what happened comes from the examination of the administrative data on display in Figure 2. Remember these suggest that the reform (i) led to an increase in the overall number of pupils retained and (ii) failed to reduce the retention rate at the end of grade 8.

Turning to PISA data, and focusing on the changes in the proportion of pupils in grade 10, we get a similar outcome. Table 2 (Equation 1) shows that the proportion of pupils in grade 10 fell by about 4 percentage points in the French Community after the 2001 reform. At this stage, we cannot be sure that this fall is indeed due to

Table 2. Determinants of the probability of attending grade 10 at the age of 15, 2003 and 2006.

Equation (1)			Equation (2)			Equation (3)		
Parameter	Estimate	*p*-value	Parameter	Estimate	*p*-value	Parameter	Estimate	*p*-value
Intercept	0.540	0.000	Intercept	0.640	0.000	Intercept	0.725	0.000
2003	0.037	0.001	2003	0.036	0.033	2003	0.040	0.033
2006	–	–	2006	–	–	2006	–	–
			Mother less than secondary	−0.181	0.000	HiseiQ20	−0.251	0.000
			Mother secondary	−0.102	0.000	HiseiQ40	−0.173	0.000
			Mother tertiary	–		HiseiQ60	−0.115	0.000
			Mother less than secondary × 20**0**3	105	0.000	HiseiQ80	−0.051	0.070
			Mother less than secondary × 2006	–		HiseiQ100	–	–
			Mother secondary × 2003	0.022	0.277	HiseiQ20 × 2003	0.057	0.001
			Mother secondary × 2006	–		HiseiQ20 × 2006	–	–
			Mother tertiary × 2003	0.036	0.033	HiseiQ40 × 2003	0.038	0.013
			Mother tertiary × 2006	–		HiseiQ40 × 2006	–	–
						HiseiQ60 × 2003	0.020	0.024
						HiseiQ60 × 2006	–	–
						HiseiQ80 × 2003	0.009	0.028
						HiseiQ80 × 2006	–	–
						HiseiQ100 × 2003	0.000	0.235
						HiseiQ100 × 2006	–	–

Source: PISA, OECD.

Note: Linear probability (OLS) estimates. French-Speaking community of Belgium. Pisa 2003 *vs*. 2006.

Equation (1) regresses grade 10 attendance on year dummy.

Equation (2) regresses grade 10 attendance on year dummy, mother's highest degree and the interaction of mother's highest degree with year dummy.

Equation (3) regresses grade 10 attendance on year dummy, the quantile of highest parental socio-economic index (HISEI (Highest occupational index score of the student's father or mother (Ganzeboom, de Graff, and Treiman 1992))) and that quantile I interacted with year dummy. Q20 means that the student has an individual HISEI score comprised between the lowest value and the 20th quantile of the overall HISEI distribution. Q40 means that he/she is between the 20th and 40th quantiles and so forth.

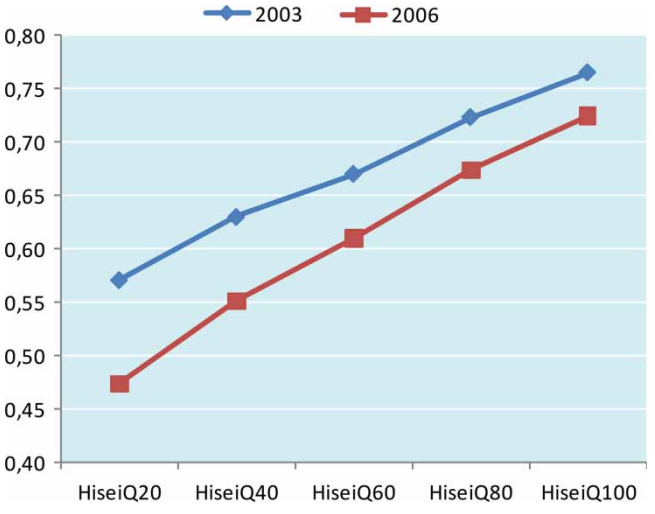

Figure 3. Making it to grade 10 at the age of 15.
Source: PISA, OECD.
Note: French-Speaking Community of Belgium. Pisa 2003 *vs*. 2006. Breakdown by quantile of HISEI. Q20, for example, means that the student has an individual HISEI score comprised between the lowest value and the 20th quantile of the overall HISEI distribution. Q40 means that he/she is between the 20th and 40th quantiles. Linear probability (OLS) estimates.

the reform. It could be that there was a negative trend in performance in the French-Speaking Community of Belgium. But, this (overall) fall in relative proportion is confirmed by a more thorough analysis that tries to assess the situation of the borderline students. These are identified first as those with a low-educated mother (less than upper secondary). Table 2 (Equation 2) shows that the fall for that group was of 14.1 percentage points between 2003 and 2006, compared with 3.6 percentage points for the students of mothers with a tertiary education attainment.

Another strategy consists of identifying borderline students by the Highest International Socioeconomic Index of Occupational Status (HISEI)[13] available in PISA (Equation (3) in Table 2 and Figure 3). The results consistently show that grade 10 attendance in 2006 is lower than that in 2003, with declines ranging from 4.9 (Q80 quantile) to 9.7 (Q20 quantile)[14] percentage points; the large reductions being observed among borderline pupils (i.e. Q20). The tentative conclusion is thus that the re-introduction of grade retention increased the proportion of pupils lagging behind at age 15, particularly among borderline ones.

4.3 *Evaluation of the threat effects of the reform on test scores*

Since the decision to retain is highly decentralised in the French-Speaking Community of Belgium – meaning that there is an inevitable element of subjectivity involved – it is not clear that the results of the previous section are fully informative about the effects of the reform on pupils' actual cognitive attainment. For example, it could be that the re-introduction of retention at grade 7 reinforced the belief among teachers that retaining pupils is a good idea and could explain rising grade retention frequencies (i.e. lower

shares of pupils making it to grade 10, see Figure 3). Thus, to investigate further the reform's impact, a second step is to evaluate its effects on PISA test scores at age 15.

Remember that due to systematic difference in the grade attended among 15-year-olds, PISA cannot help assess the *ex post*/final benefits (or costs) of grade repetition (what labour economists call the treatment effects of participation). Nonetheless, we can test for a minimal requirement for grade retention to generate (positive) threat effects, which is why it should at least have led to an improvement in average performance in grade 10. As mentioned earlier, there are two mechanisms that should drive average performance up in grade 10 under the grade retention regime: (i) a pure selection effect, due to the screening of weaker pupils, and (ii) a positive threat/motivational effect, presumably mainly on borderline pupils, thus on those who just made it to grade 10. The necessary condition for (ii) to exist is to verify that grade 10 test scores have improved after the reform.

In a nutshell, the exercise we propose consists in comparing performance pre- and post-reform conditional on being in grade 10, with a DD design. We allow for control *vs*. treated group fixed effects, common time trend and a string of time-varying socio-economic and school-level input factors. The treated cohort is the 2006 PISA cohort from the French-Speaking Community of Belgium. We estimate the following model:

$$Y_{i,t} = \theta + \beta \mathrm{BFR} + \gamma D06 + \lambda D06 \times \mathrm{BFR} + X'_{i,t}\xi + \varepsilon_{i,t}, \quad i = 1, \ldots, N,$$
$$t = 2003,\ 2006,$$

where $Y_{i,t}$ is the score of pupil i that participated to PISA during the year t; BFR is a dummy equal to 1 if pupil is from the French-Speaking Community of Belgium and 0 if pupil is from one of the European countries forming the control group[15]; $D06$ equals 0 if the observation corresponds to year 2003 (nontreated cohorts) and 1 if it was made in 2006 (treated cohorts); $X'_{i,t}$ is a vector of controls that includes the pupil's parental socio-economic background index and education attainment plus proxies of school-level spending per pupil (student/teacher ratio, share of certified teachers and the number of computers per pupil); $\epsilon_{i,t}$ is the usual random error term and N is the total number of pupils sampled by PISA for the various countries considered in the analysis.

The parameter λ should capture the effect of the reform on the average scores, and the necessary condition to validate the hypothesis of (positive) 'threat' effects corresponds to λ being statistically significantly superior to 0.

Table 3 presents the results of the DD estimates (estimates of λ), using the individual scores in Math, Science or Reading literacy (or their simple average). We first estimate λ using all grade 10 pupils of the French-Speaking Community of Belgium (Table 3, Equation 1). We complement these by several re-estimations of Equation (1) using -quantile regression and focusing on the lower quantiles of the grade 10 score distribution. The aim is to increase the (relative) weight of borderline students in the comparison, assuming that the students who were the most likely to respond to enhanced threat concentrate at the bottom of the grade 10 distribution. As done in the previous section, we also isolate those (i) with a low-educated mother (Table 3, Equation 2) and (ii) belonging to the lowest quantiles of the HISEI[16] distribution (Table 3, Equation 3).

The results in Table 3 are not supportive of the threat–benefit hypothesis. We fail to find positive λ's when focusing on Q50 (i.e. median) for the mean score (Equation 1,

Table 3. DD[a] estimates (λ), grade 10 test scores and focus on various quantiles[b] of the score distribution (Equation 1), on pupils with low-educated mothers[c] (Equation 2) or below the 40th percentile of the HISEI[d] distribution (Equation 3), 2003 and 2006.

		Equation (1)		Equation (2)		Equation (3)	
Dependent	Quantiles	Estimate	*p*-value	Estimate	*p*-value	Estimate	*p*-value
Mean score[e]	ALL/Q50	1.79	0.576	10.77	0.081	-0.50	0.919
	Q40	4.34	0.181				
	Q30	3.91	0.233				
	Q20	6.20	0.126				
Reading	ALL/Q50	6.40	0.041	16.25	0.015	4.27	0.419
	Q40	4.91	0.123				
	Q30	7.24	0.052				
	Q20	7.54	0.065				
Science	ALL/Q50	8.66	0.009	17.42	0.009	9.02	0.089
	Q40	8.31	0.035				
	Q30	8.26	0.018				
	Q20	10.71	0.009				
Math	ALL/Q50	−4.01	0.240	−1.36	0.829	−14.77	0.003
	Q40	−8.29	0.013				
	Q30	−6.97	0.108				
	Q20	−6.82	0.106				

Source: PISA, OECD.

[a]All observations are weighted by the (inverted) country-specific PISA sample size. This is a way to reduce the risk that big countries – which sampled more pupils to achieve national representativeness – dwarf smaller ones. All regressions include the following controls: father and mother socio-economic background index, father and mother education attainment, school-level student/teacher ratio, school-level share of certified teachers and school-level computer/student ratios.

[b]Using quantile regression.

[c]Mothers without an upper secondary degree.

[d]HISEI corresponds to the highest occupational index score of the student's father or mother (Ganzeboom, de Graff, and Treiman 1992).

[e]Student's average test score in Math, Sciences and Reading.

upper rows). When we concentrate on the pupils belonging to the lower quantiles (Q40, Q30 and Q20) of the mean score distribution (lower rows), estimated λ's rise a bit, but remain not statistically different from zero. The breakdown by topic/field delivers mixed evidence. Reading and Science scores seem to have improved slightly, whereas Maths scores have deteriorated (though not statistically significantly). What is more, coefficients associated with the lower quantiles do not deviate significantly from those obtained with Q50.

The focus on pupils with low-educated mothers (Equation 2) also delivers a mixed message. There is no improvement in Maths, but some (statistically significant at the 5% level) in Science and Reading. The examination of pupils below the 40th percentile of the HISEI distribution (Equation 3) suggests an absence of significant change in Reading and Science, but a significant deterioration of attainment in Maths. These results (together with those discussed in Section 4.2) provide no strong evidence supportive of any kind of positive threat effects of grade retention.

5. Conclusion

This paper exploits a 2001 reform in the French-Speaking Community of Belgium (re)introducing the possibility to impose grade retention at the end of grade 7 (first grade of secondary education). It does so with the aim of evaluating the *threat effects* of grade retention. Contrary to participation (or treatment) effects, threat effects of grade repetition sanction have received little attention in the literature. Another plus is that the Belgian reform constitutes a 'natural experiment', introducing an exogenous variation in the level of grade repetition threat.

The main results are that the Belgian experiment did not lead to better outcomes, even among the segments of the borderline population of students. The first measure of educational performance we consider is the position in the curriculum (or grade). To make sure that the effect we identify is indeed due to the reform, we distinguish pupils according to their likelihood of having been affected by the reform (i.e. borderline *vs*. other students). We show that the overall proportion of pupils who made it to grade 10 (the reference grade) at the age of 15 fell by about 4 percentage points. The corresponding fall among more borderline pupils is estimated to be 10–14 percentage points.

We consider a second measure of outcome: grade 10 pupils test scores in Maths, Science and Reading literacy measured by PISA. Changes in grade 10 test scores are likely to reflect a (presumably positive) *screen out effect* due to a more selective access to grade 10. But, it should also capture the *threat effects* singularly among the borderline students, forming the lower end of the grade 10 score distribution. Both effects should *a priori* lead to an improvement in the average test scores in grade 10. But, we 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. This result contrasts with those regularly obtained by the literature that evaluates the threat effects of ALMPs.

Notes

1. As opposed to what he calls the 'sanction' effect (i.e. higher achievement that could be ascribed to grade repetition and the additional time granted to students to master the lower grade curriculum, before moving on).
2. Students who do not meet the standards are required to attend a 6-week summer school programme, after which they retake the exams. Those who pass move on to the next grade; those who fail this second exam are required to repeat the grade (Jacob 2005).
3. Belgium is a federal state where the educational policy is split according to linguistic lines. Each linguistic community is in charge of its educational system. Only minor aspects of the educational policy (like the age of compulsory education, i.e. 18) remain under federal jurisdiction.
4. *Décret relatif à l'organisation du premier degré de l'enseignement secondaire* D. 19 July 2001. M.B. 23 August 2001.
5. Formally, the legislator insists on the fact that the reform's aim was not exactly to force the pupils to 'repeat' the year, but to channel weaker students (who do not achieve satisfactory results at the end of grade 7 or at the end of grade 8) towards a 'complementary' year. In practice, however, it amounts to imposing that these students take more time before moving to the upper grade.
6. Remember that we look at age 15 scores to identify the effect of a decision that affected pupils when they were aged 12–13.
7. We have chosen not to use the 2000 wave because the sample size for the French Community of Belgium was about half the size in comparison to 2003 and 2006. This may raise issues of comparability across cohorts.

8. Following what is usually done in the empirical labour literature on the threat benefits of ALMPs.
9. We essentially tried alternatives using the Flemish Community of Belgium and the immediate EU neighbours of Belgium (France, Germany and the Netherlands). But, the choice of countries forming the control group did not fundamentally affect our results.
10. We aggregate these different countries, weighing each of them by the inverse of its PISA sample size. More details on this are given in the next section.
11. Scores reported in Table 1 and in the subsequent econometric analysis correspond to individual averages, aggregating scores obtained in the three topics covered by PISA: Math, Sciences and Reading literacy.
12. PISA test scores are based on standardised questionnaires that teams of experts have assessed as to their capacity to gauge pupils' skills and competences and make them comparable across waves and across participating countries. By contrast, in the French-Speaking Belgium, the retention decision is based on the teachers' assessment of the pupil's ability of passing to a higher grade. There is no standardised test used across schools, nor is there a clearly defined threshold to determine whether a pupil should be retained or not.
13. Which corresponds to the highest occupational index score of the student's father or mother (Ganzeboom, de Graff, and Treiman 1992).
14. Q20 means that the student has an individual HISEI score comprised between the lowest value and the 20th quantile of the overall HISEI distribution. Q40 means that he/she is between the 20th and 40th quantiles.
15. Great Britain, Greece, Hungary, Ireland, Iceland, Norway, Poland and Sweden, all European countries where grade retention does not exist.
16. Highest occupational index score of the student's father or mother.

References

Black, D.A., J.A. Smith, M.C. Berger, and J.N. Brett. 2003. Is the threat of reemployment services more effective than the services themselves? Evidence from random assignment in the UI system. *American Economic Review* 93, no. 4: 1313–27.

Cockx, B., and M. Dejemeppe. 2007. Is the notification of monitoring a threat to the unemployed? A regression discontinuity approach. IZA Discussion Paper no. 2854. Bonn: Institute for the Study of Labor.

Eide, E.R., and M.H. Showlater. 2001. The effect of grade retention on educational and labour market outcomes. *Economics of Education Review* 20, no. 6: 563–76.

Ganzeboom, H.B.G., P. de Graff, and D. Treiman. 1992. A standard international socio-economic index of occupational status. *Social Science Research* 21, no. 1: 1–56.

Geerdsen, L.P. 2006. Is there a threat effect of labour market programmes? A study of ALMP in the Danish UI system. *The Economic Journal* 116, no. 513: 738–50.

Geerdsen, L.P., and A. Holm. 2007. Duration of UI periods and the perceived threat effect from labour market programmes. *Labour Economics* 14, no. 4: 639–52.

Goos, M., J. Van Damme, P. Onghena, and K. Petry. 2010. First-grade retention: Effects on children's academic and psychosocial growth throughout primary education. Report to the Flemish Minister of Education, Programme 'Steunpunten voor Beleidsrelevant Onderzoek', Brussels.

Grissom, J., and L. Shepard. 1989. Repeating and dropping out of school. *Flunking grades: Research and policies on retention*, ed. L. Shepard and M. Smith, 34–63. London: The Palmer Press.

Heckman, J., R. LaLonde, and J. Smith. 1999. The economics and econometrics of active labor market programs. *Handbook of labor economics*, ed. O. Ashenfelter and D. Card, chap. 31, vol. IV, 1865–2073. Amsterdam: Elsevier.

Holmes, C.T. 1989. Grade level retention effects: A meta-analysis of research studies. *Flunking grades: Research and policies on retention*, ed. L.A. Shepard and M.L. Smith, 16–33. London: Falmer.

Jacob, B.A. 2005. Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago public schools. *Journal of Public Economics* 89, no. 5–6: 761–96.

Jacob, B.A., and L. Lefgren. 2004. Remedial education and student achievement: A regression-discontinuity analysis. *Review of Economics and Statistics* LXXXVI, no. 1: 226–44.
Jacob, B.A., and L. Lefgren. 2009. The effect of grade retention on high school completion. *American Economic Journal: Applied Economics, American Economic Association* 1, no. 3: 33–58.
Jimerson, S.R. 1999. On the failure of failure. *Journal of School Psychology* 37, no. 3: 243–72.
Jimerson, S.R. 2001. Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review* 30, no. 3: 420–37.
Manacorda, M. forthcoming. The cost of grade retention. *Review of Economics and Statistics,* doi:10.1162/REST_a_00165.
Roderick, M. 1994. Grade retention and school dropout: Investigating the association. *American Educational Research Journal* 31, no. 4: 729–59.
Roderick, M., and J. Nagaoka. 2005. Retention under Chicago's high-stakes testing program: Helpful, harmful, or harmless? *Educational Evaluation and Policy Analysis* 27, no. 4: 309–40.
Rosholm, M., and M. Svarer. 2004. Estimating the threat effect of active labour market programmes. IZA Discussion Paper no. 1300. Bonn: Institute for the Study of Labor (IZA).
Rosholm, M., and M. Svarer. 2008. The threat effect of active labour market programmes. *Scandinavian Journal of Economics* 110, no. 2: 385–401.