9

Evaluating the "threat" effects of grade repetition exploiting the 2001 reform by the French-speaking community of Belgium

Michèle Belot

Oxford University, Nuffield Centre for Experimental Social Sciences (CESS),
michele.belot@nuffield.ox.ac.uk

Vincent Vandenberghe (corresponding author)

Université catholique de Louvain (UCL), Institut de Recherches Economiques et Sociales (IRES)
vincent.vandenberghe@uclouvain.be

Evaluating the "threat" effects of grade repetition exploiting the 2001 reform by the French-speaking community of Belgium

Michèle Belot

Oxford University, Nuffield Centre for Experimental Social Sciences (CESS),
michele.belot@nuffield.ox.ac.uk

Vincent Vandenberghe (corresponding author)

Université catholique de Louvain (UCL), Institut de Recherches Economiques et Sociales (IRES)
vincent.vandenberghe@uclouvain.be

Like active labour market programmes (ALPMs), 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. That year, policy makers reinstated the possibility to repeat grade 7, putting an end to the regime of "social promotion" that existed before. We use data from two waves of the PISA study (corresponding to periods before and after the reform) to evaluate the medium-term effects of the 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 outcomes, even among the segments of the population the most at risk of grade repetition.

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 rates across OECD countries (Figure 1). Countries/entities like 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 like Norway (NOR), Sweden (SWE), Japan (JPN) and Great Britain (GBR) have no grade retention at all.

Figure 1 – Average score in math and share of pupils aged 15 attending reference grade[a].

Year 2006

Share of pupils attending reference grade
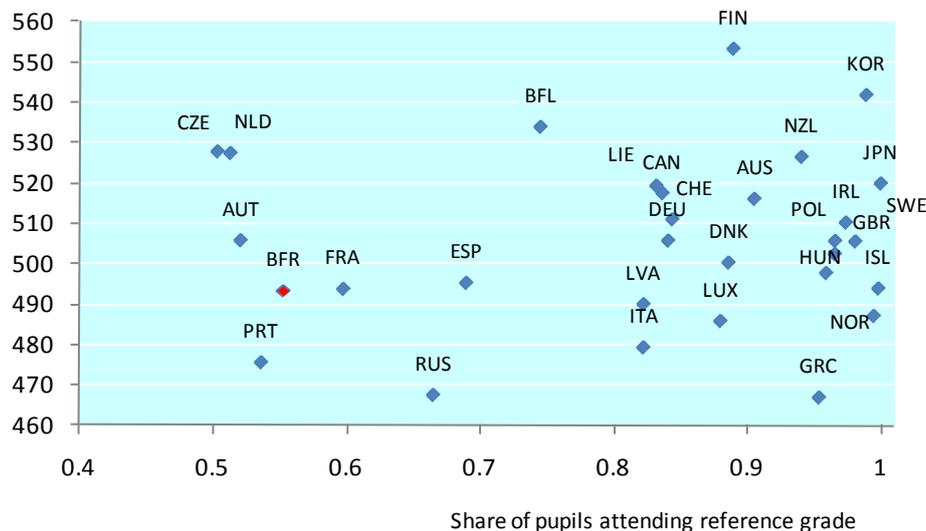
a: Grade 10 in most countries, grade 9 otherwise. The grade of reference is identified as the most attended grade among 15 year-olds who participated to PISA.
ARG: Argentina ; AUS: Australia ; AUT: Austria ; AZE: Azerbaijan ; BFR: French-Speaking Community of Belgium; BFL: Flemish-Speaking Community 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: 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.
Source: PISA 2006

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 consists 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 but 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 year, policy makers reinstated the possibility to repeat grade 7 (1st year of secondary education), putting an end to the regime of "social promotion" that existed before.

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 into 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 a priori more exposed to the risk of grade repetition) is estimated to be in the range of 10-11 percentage points.

We consider a second measure of outcome: test scores. These are reflecting the actual cognitive skills attained by pupils at a certain grade. And 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 of 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 epitomizes 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…). These seem to suffice to entice unemployed individuals to raise their search effort prior to participation and rapidly move out of unemployment (Roshomm & Svaer, 2004; Cockx & Dejemeppe, 2007; Rosholm and Svarer, 2008). 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 have small, and in some cases even adverse, effects in terms of increasing job-finding rates for the long-term unemployed; see *e.g.* 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 to 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). 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 benefit by both lowering dropout rates and raising labour market earnings, although their 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 1996-97. The policy created a discontinuity in the relation between scores in a single standardised test (thereby the label "high stakes") 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 two 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 (2008) exploits a discontinuity induced by a rule establishing automatic grade retention for pupils missing more than 25 days and shows that grade retention leads to a substantial increase in drop-out and lower educational attainment even 4 or 5 years later.

In short, these studies nicely evaluate the causal effects of being retained. Like many ALMPs evaluation studies they 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 Roderick and Engel (2001) is a noteworthy exception. Using survey responses of pupils affected by the above-mentioned Chicago high-stake policy, they document some evidence of motivational effects of such a policy. The majority of these students described increased *work effort* under the policy. They reported greater attention to class work or more time spent studying outside school. They also find that students with high levels of work effort generally had greater-than-average learning gains and positive promotional outcomes.

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 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 Belgium, and is particularly frequent in the French-Speaking Community[1] (Figure 1). The retention decision in French-Speaking Belgium 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

---

[1] 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.

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 reducing its incidence at the end of grade 7 (1st year of secondary education). The argument for suppressing grade retention 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 no grade retention was allowed at the end of grade 7, a decision that translated into a sharp fall of the number of "repeaters" (Figure 2). During that period, grade retention sanctions could only be pronounced at the end of grade 8. Pupils could only possibly repeat grade 7 upon agreement between parents and teachers. This is why on Figure 2 one 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 six years later. In September 2001, the decision was made[2] 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 2001-02 it became possible to repeat grade 7 or grade 8, although not both.[3] 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 as high as before the reform; those who passed grade 7 were exposed to the threat of having to repeat grade 8 in the same way 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 (grade 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 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.

---

[2] *Décret relatif à l'organisation du premier degré de l'enseignement secondaire* D. 19-07-2001 M.B. 23-08-2001

[3] 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.

Figure 2 – Incidence of grade retention at grade7 and grade 8. School year 1992-93 to 2003-04
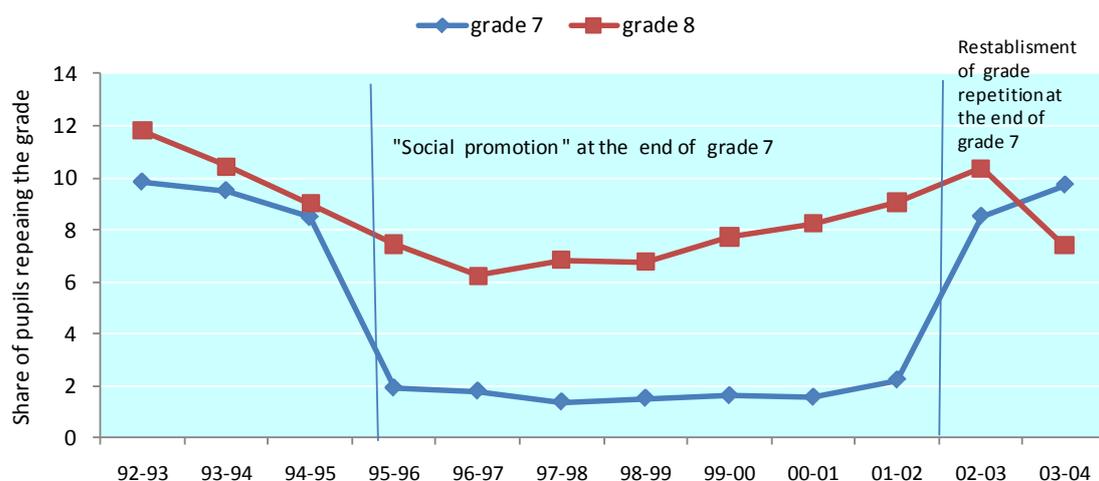
Source: French-Speaking Community of Belgium, Ministry of Education.

Administrative data (Figure 2) show that the number of pupils repeating grade 7 rose from the school year 2002-03 onwards. Quite strikingly, the same data also show that the total number of students repeating grade 7 *or* grade 8 is substantially higher in 2002-03 and beyond, 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): the OECD's international standardised assessment administered to 15-year-olds students. This allow us to investigate more carefully the medium- to long-term[4] (causal) threat effects of the 2001 reform on *i)* grade attainment and *ii)* standardized test scores.

Note already that PISA only contain *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. non- repeaters) would be informative about whether repletion is beneficial or not. Since in PISA there are differences in positions in the curriculum, the comparison of test scores between retained and non-retained pupils is not directly meaningful. By contrast, PISA test scores are perfectly suitable to assess the threat effects of grade repetition.

---

[4] Remember that we look at age 15 scores to identify the effect of a decision that affected pupils when they were aged 12-13.

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 these three topics). In addition, the PISA assessment contains many questions about attended grade and programme student background or school characteristics.

We use two waves of the PISA study (2003 and 2006). Teh 2003 wave includes pupils who were unaffected by the reform (no risk of retention in grade 7) and the 2006 wave includes pupils who were affected by the reform.[5] Since we are looking at 15-year olds, the grade reference for these pupils is grade 10 (grade they would have attained without retention).

The identification of a causal effect of grade retention threats faces challenges. First, 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. Second, 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 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*.

Of course, the challenge is to find an adequate control group. There is no ideal counterfactual country. We have opted for a group of countries[6] *i) located in Europe ii) where grade retention does not exist*. This group comprises Great-Britain, Greece, Hungary, Ireland, Iceland, Norway, Poland and Sweden. Why Europe? Because the "parallel-trend" assumption is a critical assumption in a DD framework. An non-European OECD countries could be affected by overall trends that are not similar to those affecting Europe.[7]

---

[5] 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.

[6] We aggregate these different countries, weighing each of them by the inverse of its PISA sample size. More details on this in the next section.

[7] The US, 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 that 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[8]. 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 the very aggregate level. The third column of Table 1 for instance suggests that overall test scores[9] gaps remained within the range of 2 to 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. Socioeconomic characteristics influencing PISA scores could also have diverged between 2003 and 2006. For instance, the shares of students with low-educated mother may have evolved In a diverging manner. Yet, Table 1 descriptive statistics on key aspects of pupils' socioeconomic background do not support this. What is more, we are able to account for socioeconomic 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* reinforce the plausibility that 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 Belgium with just one other country) simply cancel out in the *one-to-many* setting used here.

In short, our study presents five main advantages. First, it examines the *threat* effects of grade retention; something that has received too little attention so far in the literature, although the threat/incentive argument lies at the core of the discussions and policy debates surrounding grade repetition. Second the reintroduction of grade retention in 2001 provides a *natural experiment* to evaluate the effects of grade retention threat. Of course the Belgian reform only suppressed grade retention at the end of grade 7, so it is not clear that the 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. Third, we evaluate the *medium-term effects* of the reform. We have information from the PISA study (before the reform (PISA 2003) and after the reform (PISA 2006), measuring, at the age of 15, grade attainment and cognitive performance in a standardised test across OECD countries in Maths, Reading and Science. Pupils who have not repeated a year should

---

[8] 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.

[9] Scores reported in Table 1 and in subsequent econometric analysis correspond to individual averages, aggregating scores obtained in the three topics covered by PISA: math, sciences and reading literacy.

then be in grade 10, thus three grades further than the one affected by the reform. Since schooling is compulsory until the age of 18 in Belgium, there is no possible drop-out yet. This facilitates the comparison of scores pre and post-reform. Fourth, we compare performance across two *different "threat" regimes*, rather than within regime. Firth, our PISA data provide a *control group* suitable for carrying out at difference-in-differences (DD) analysis, namely the European countries that do not have a grade repetition policy.

Table 1. **French-Speaking Belgium vs. European controls. Descriptive statistics before (2003) and after (2006) natural experiment**

Entity	Year	Nobs	Average score[a]	Student /teacher	Computer per student	Certified tearchers (%)	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	39228	496.68	11.43	0.20	0.94	49.58	3.39	3.34
	2006	45571	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

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 student 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 non-tertiary education, level 5 – First stage of tertiary education, level 6 – Second stage of tertiary education.
Source: PISA 2003, 2006

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 reintroduced at all grades. The reduction is even more significant among borderline pupils.

Our second measure of outcome are grade 10 test scores. These reflect the cognitive skills attained by pupils; a more objective measure of attainment.[10] We can test for the necessary condition for an enhanced threat of grade retention to have generated benefits, which is that it should be associated with an improvement in the distribution of scores. There are two possible mechanisms which could play a role: (1) a *screen out effect*: grade retention effectively prevents weaker students from gaining access to grade 10 at the age of 15, (2) a *threat/incentive effect*: the threat of having to repeat a year should stimulate the pupils "at risk" to work harder. Both

---

[10] 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.

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 rather than just at the end of grade 8 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 happed comes from the examination of the administrative data on display on Figure 2. Remember these suggest that the reform *i)* led to an increase of 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 outcomes. Table 2 (equ. 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 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 (equ. 2) shows that the fall for that group was of 10 percentage points between 2003 and 2006.

Another strategy consists of identifying borderline students by the Highest International Socioeconomic Index of Occupational Status (*Hisei*)[11] available in PISA (Table 2, equ. 3 & Figure 3). The results consistently show that grade 10 attendance in 2006 is lower than in 2003, with declines ranging from 4.9 (upper Q80 quantile) to 9.8 (lowest Q20 quantile) percentage points; the large reductions being observed among borderline pupils (i.e. Q80). 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.

---

[11] Which corresponds to the highest occupational index score of the student's father or mother (Ganzeboom, *et al.*, 1992).

Table 2 – Making it to grade 10 at the age of 15. French-Speaking Community of Belgium.

Pisa 2003 vs. 2006

Equation 1			Equation 2			Equation 2		
Parameter (2006=ref)	Estimate	Probt	Parameter (2006=ref)	Estimate	Probt	Parameter (2006, Hisei Q100=ref)	Estimate	Probt
Intercept	0,540	0,0000	Intercept	0,640	0,0000	Intercept	0,725	0,0000
2003	0,037	0,0011	Mother less than second.	-0,181	0,0000	2003	0,040	0,2351
			Mother second.	-0,102	0,0000	Hisei[a] Q20	-0,251	0,0000
			Mother tertiary	0,000		Hisei[a] Q40	-0,173	0,0000
			Mother less than second. *2003	0,105	0,0002	Hisei[a] Q60	-0,115	0,0001
			Mother second. *2003	0,022	0,2774	Hisei[a] Q80	-0,051	0,0699
			Mother tertiary *2003	0,036	0,0326	Hisei[a] Q20*2003	0,098	0,0007
						Hisei[a] Q40*2003	0,079	0,0134
						Hisei[a] Q60*2003	0,061	0,0239
						Hisei[a] Q80*2003	0,049	0,0278

a:Highest <u>occupational</u> index score of the student's father or mother (Ganzeboom, *et al.*, 1992).
Source: PISA 2003, 2006

Figure 3 – Making it to grade 10 at the age of 15. French-Speaking Community of Belgium.

Pisa 2003 vs 2006. Breakdown by quantile of Highest International Socioeconomic Index of

Occupational Status (*Hisei*). OLS estimates.
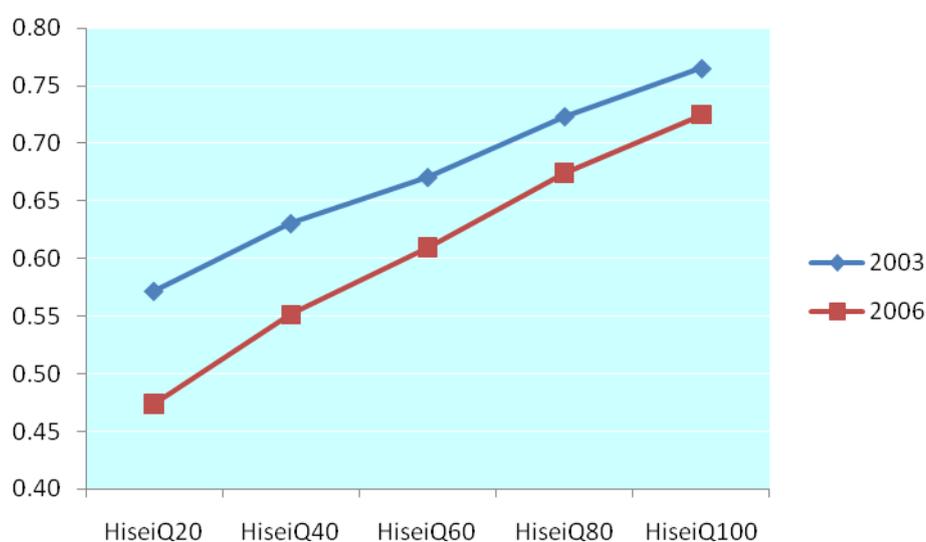
Source: PISA 2003, 2006

4.3   Evaluation the threat effects of the reform on test scores

Since the decision to retain is highly decentralized in the French-Speaking Community of Belgium - meaning 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 that it should at least succeed have lead 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 a *ii)* positive threat/motivational effect, presumably mainly on borderline pupils, thus on those who just made it into grade 10. And 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 socioeconomic 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:

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

where

- - $Y_{i,t}$ is the score of pupil *i* that participated to PISA during the year *t*;
- - country is a vector of country[12] dummies, BFR is a dummy equal to 1 it pupil is from the French-Speaking Community of Belgium and 0 if is she is from one of the European countries forming the control group.[13]

---

[12] Other than the French-Speaking Community of Belgium.
[13] Great-Britain, Greece, Hungary, Ireland, Iceland, Norway, Poland and Sweden, all European countries where grade retention does not exist.

- • - *D06*, equal to 0 if the observation corresponds to year 2003 (non-treated cohorts) and 1 if it was made in 2006 (treated cohorts);

- • - $X'_{i,t}$ is a vector of controls that include the pupil' 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 computer per pupil);

- • - $\varepsilon_{i,t}$ the usual random error term;

- • - *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 distinguish between groups of pupils as we expect to see the largest changes in the distribution of scores for the groups that have experienced the largest change in the proportion of pupils in grade 10 (our borderline pupils). We first estimate λ using all grade 10 pupils of the French-Speaking Community of Belgium (Table 3, equ. 1*)*. We complement these by several re-estimations of equation 1 where we incrementally eliminating the upper percentiles of overall distribution of scores (Figure 3). The aim is still 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 were 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, equ. 2) ,and *ii)* those belonging to the lowest quantiles of the *Hisei*[14] distribution (Table 3, equ. 3).

The results on display in Table 3 are not supportive of the threat benefit hypothesis. We fail to find positive λs when using the whole grade 10 sample (equ. 1, upper rows). When we concentrate on the pupils belonging to the lower quantiles of the math, science or reading score distribution (lower rows), we also find no evidence of statistically *significant positive λs*.

The focus on pupils with low-educated mothers (equ. 2) 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 (equ. 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 evidence supportive of any kind of positive threat effects of grade retention...

---

[14] Highest occupational index score of the student's father or mother.

Figure 3 – Increasing the chances of identifying borderline individuals among grade 10 students: eliminating the upper end part of score distribution.

Grade 10

<pth perc

Pisa Score
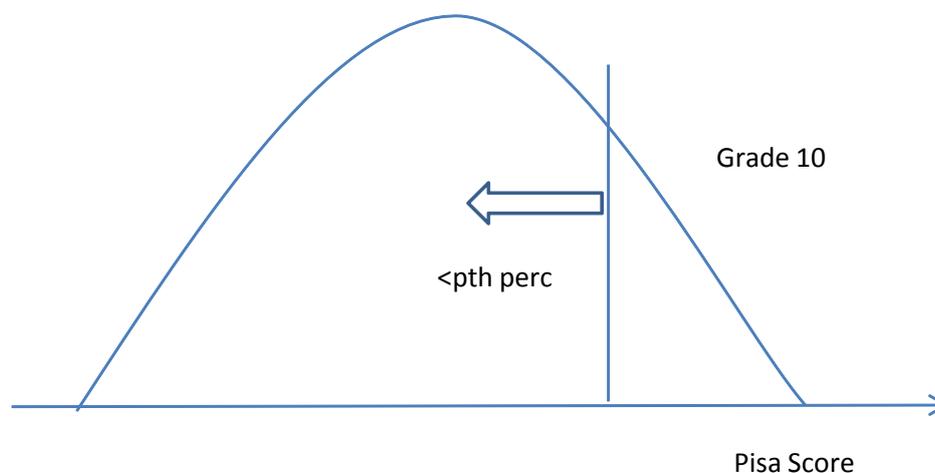
5   Conclusion

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 (1st grade of secondary education). It does so with the aim of evaluating the *threat effects* of grade retention. Contrary to the *ex post* (or participation (or treatment) effects, the 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 the level of grade repetition threat.

The main results are 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 overall proportion of pupils who made it into 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 10-11 percentage points.

We consider a second measure of outcome: grade 10 pupils test scores in Maths, Science and Reading literacy measured by PISA. Changes of 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 active labour market programmes (ALMPs).

Table 3 – Difference-in-Differences[a] estimates (λ). Grade 10 test scores. Focus on various quantiles[b] of the score distribution (equ. 1), on pupils with low-educated mother[c] (equ. 2) or below the 40th percentile of the *Hisei*[d] distribution (equ.3).

Topic	Retained quantile	Equation 1		Equation 2		Equation 3	
		Estimate	Probt	Estimate	Probt	Estimate	Probt
Mean score[e]	**all**	2,609	0,7489	15,058	0,0000	-5,146	0,1321
	l90	0,492	0,9457				
	l80	0,806	0,8909				
	l70	-0,855	0,8915				
	l60	0,806	0,8909				
	l50	0,798	0,8841				
Math	**all**	-4,402	0,6204	4,564	0,1932	-18,572	0,0000
	l90	-5,773	0,4492				
	l80	-8,389	0,2284				
	l70	-8,695	0,1798				
	l60	-6,905	0,2422				
	l50	-8,699	0,1168				
Science	**all**	7,518	0,4010	21,515	0,0000	4,461	0,2413
	l90	4,137	0,5977				
	l80	2,021	0,7810				
	l70	3,029	0,6572				
	l60	4,562	0,4778				
	l50	6,217	0,3006				
Reading	**all**	4,796	0,5745	19,094	0,0000	-1,329	0,7170
	l90	1,951	0,7994				
	l80	2,152	0,7669				
	l70	3,606	0,6002				
	l60	3,048	0,6447				
	l50	3,914	0,5466				

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: Characterising the overall (all grades pooled) test score distribution.
c: Mothers without an upper secondary degree
d: Hisei corresponds to highest occupational index score of the student's father or mother (Ganzeboom, et al., 1992)
e: Student's average test score in math, sciences and reading
Source: PISA 2003, 2006

6   References

Black, D. A., Smith, J. A., Berger, M. C. and Brett J. N. (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(4), pp. 1313-1327.

Bart Cockx & Muriel Dejemeppe (2007), "Is the Notification of Monitoring a Threat to the Unemployed? A Regression Discontinuity Approach," *IZA Discussion Papers* 2854, Institute for the Study of Labor.

Eide, E.R. and M.H. Schowlater (2001), The Effect of Grade Retention on Educational and Labour Market Outcomes, *Economics of Education Review*, 20(6), pp. 563-576.

Ganzeboom, HBG., de Graff, P. and Treiman, D. (1992), "A Standard International Socio-Economic Index of occupational status", *Social Science Research*, 21, pp. 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, pp.738–750.

Geerdsen, L.P, & A. Holm (2007), Duration of UI periods and the perceived threat effect from labour market programmes, *Labour Economics*, 14, pp. 639-652.

Grissom, J. and L. Shepard (1989), *Repeating and dropping out of school*, in L. Shepard and M. Smith (Eds.), Flunking Grades: Research and policies on Retention, London: The Palmer Press.

Heckman, J., R. LaLonde and J. Smith (1999), "The Economics and Econometrics of Active Labor Market Programs," in O. Ashenfelter and D. Card, Chapter 31, *Handbook of Labor Economics* , Vol. IV, 1865-2073.

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

Jacob, B.A & L. Lefgren, (2004), Remedial Education and Student Achievement: A Regression-Discontinuity Analysis, *Review of Economics and Statistics,* LXXXVI(1), pp. 226-244.

Jacob, B.A & L. Lefgren, (2009), The Effect of Grade Retention on High School Completion, *American Economic Journal: Applied Economics*, American Economic Association, 1(3), pp. 33-58.

Jimerson, S.R. (1999), On the Failure of Failure, *Journal of School Psychology*,37(3), pp. 243–272.

Jimerson, S.R. (2001), Meta-analysis of Grade Retention Research: Implications for Practice in the 21st Century, *School Psychology Review* 30(3), pp. 420-437

Manacorda, M. (2008), The Cost of Grade Retention, *CEP Discussion Papers* 0878, Centre for Economic Performance, LSE.

McCoy, A.R. & A. J. Reynolds (1998), Grade Retention and School Performance: An Extended Investigation, *Institute for Research on Poverty, Discussion Paper no. 1167-98,* University of Wisconsin,-Madisson.

Roderick, M. (1994), Grade retention and school dropout: Investigating the association. *American Educational Research Journal*, 31, pp. 729–759.

Rosholm, Michael & Svarer, Michael (2004),Estimating the Threat Effect of Active Labour Market Programmes, *IZA Discussion Papers* 1300, Institute for the Study of Labor (IZA).

Rosholm, Michael & Svarer, Michael (2008), The Threat Effect of Active Labour Market Programmes, *Scandinavian Journal of Economics*, 110(2), pp. 385-401

Roderick, M. and M. Engle (2001), The grasshopper and the ant: motivational responses of low-achieving students to high-stakes testing, *Educational Evaluation and Policy Analysis*, 23(3), pp. 197-227.

Roderick, M., & Nagaoka, J. (2005). Retention under Chicago's high-stakes testing program: Helpful, harmful, or harmless? *Educational Evaluation and Policy Analysis*, 27(4), pp. 309-340.