Children left behind

New evidence on the (adverse) impact of grade retention on educational careers

Dalit Contini
University of Torino

Guido Salza
University of Trento

This paper analyses the causal effect of grade retention on students' high school track decisions in Italy. Using longitudinal administrative data, we propose a matching strategy to assess the impact of retention in institutional settings with substantial leeway in the promotion/retention decision. If decision makers differ in their propensity to retain students (strictness), it is reasonable to expect some of the students kept back to be fully on a par with others who are promoted. Assuming that strictness is a school-related factor, we argue that it is not theoretically possible to find a good match within the same school. We therefore match retained students to students with a similar array of ability measures, who were promoted in schools displaying a lower degree of strictness. We find that grade retention dramatically increases school dropout rates and that the negative impact is stronger for students with poorly educated or immigrant parents.

1. Introduction

Grade retention, the practice of holding back low-performing students to repeat a grade, is a controversial issue that continues to generate much debate within and outside academia. Retention is mainly used as a remedial measure, and many experts argue that students with learning difficulties should be protected from overly demanding curricula and given extra time to catch up and develop emotional and cognitive maturity before advancing to the next grade (Crahay et al. 2008; Shepard and Smith 1989). Teachers, principals, and parents often welcome grade retention as an incentive to elicit effort from demotivated students (Belot and Vandenberghe 2014) and as a means of safeguarding quality standards in schools (Brophy 2006). However, the OECD and other international institutions have called for limiting the use of grade repetition, considering it to be a waste of human capital at the system level (OECD 2019; UNESCO 2012). They recommend setting differentiated learning standards instead, arguing that contemporary education systems must accommodate a wide range of audiences, whose members vary in educational background, needs and aspirations. Other arguments against grade retention involve the issue of stigma attached to school failure, which can lead to demotivation and loss of self-esteem and generate psychological costs that can further hinder school engagement (Shepard and Smith 1989).

Research on the impact of grade retention on students' educational careers has produced inconsistent results. One major challenge to studies of this sort is that they require detailed longitudinal data on individual careers, with relevant information about pupils' performance both before they were held back, to allow meaningful like-with-like comparisons, and afterwards, to allow analysis of the consequences on different educational outcomes. Such data is generally difficult to obtain, and often necessitates *ad hoc* integration of data archives. The main challenge to making appropriate comparisons between treated units (retained pupils) and control units (promoted pupils) is that the former generally perform worse than the latter. By exploiting the peculiarities of different institutional contexts, a variety of empirical strategies have been employed in the literature to address this issue, ranging from regression discontinuity design, instrumental-variable estimation, and matching strategies.

Apart from the informational and methodological challenges, another hurdle to this research strand is the limited generalizability of findings across institutional contexts. This is because the practice and timing of grade retention differs across educational systems. Most of the existing research come from the U.S., where grade retention is common only in primary school, whereas in Europe, it is more commonly seen in high school. Moreover, in some systems, the decision to hold students back is based on standardized tests, while in others, it is subject to a considerable degree of discretion on the part of teachers and principals.

In this paper, we analyze the impact of grade retention in Italian high schools, where retention remains largely understudied despite its widespread practice. Although retention crops up periodically in public discourse, there have been no attempts by scholars and educational authorities to pursue systematic empirical analysis of its impact. A possible explanation for is the lack of adequate data for satisfactorily addressing the question.

Here we draw on a rich longitudinal database that was constructed specifically for the purposes of this study. The dataset covers the entire population of high school students for school years 2013-14 through 2016-17 in the three most populated regions in northern Italy (Lombardy, Piedmont, and Veneto). It is based on the National Register of Students, which follows pupils over time and records all the relevant details of their educational paths – grades, school transfers, promotion or retention at the end of each school year – while simultaneously keeping track of the school and class identifiers. This data was integrated with individual-level information provided by the National Institute for the Evaluation of the Educational System (INVALSI), which administers yearly standardized tests of reading and math literacy to students in different grades and collects information on their socioeconomic and migration background.

Our contribution to the literature is also a methodological one. Exploiting the richness of the data, we propose a matching strategy to assess the causal impact of retention in institutional contexts that allows ample leeway in deciding whether or not to promote students. Because of this leeway, if decision makers differ in their tendency to hold students back – which we refer to as 'strictness' – one should be able to find fully comparable retained and promoted students. Assuming that strictness is defined at the school level, we argue that it is not theoretically possible to find a good match within the same school. Therefore, we match students who were held back to students with a similar array of ability measures who were promoted in schools displaying a lower degree of strictness.

Our findings indicate that being exposed to grade retention sharply raises the probability to drop out or transfer to less demanding school tracks. We conclude that grade retention is an inadequate remedial policy, as it ends up increasing early dropout rates and can affect the number of young people who graduate from university. Apart from its ineffectiveness, retention also creates problems of equity, as students from disadvantaged backgrounds are found to experience stronger negative consequences. Moreover, related work with the same data archive shows that, *ceteris paribus*, socially disadvantaged students face a higher risk of being held back (Salza 2022). Overall, grade retention appears to exacerbate the level of educational inequalities across social groups.

The remainder of this paper is organized as follows. We begin with an overview of the recent literature on the impact of grade repetition across countries (Section 2). Section 3 provides an outline of the Italian institutional context and the rules governing grade progression and Section 4 describes the data. In

Section 5, we discuss the identification strategy and describe the matching procedure and in Section 6 we present our findings. At the end of the paper, we discuss the results and draw conclusions.

2. Previous literature

Scholars have long debated the impact of grade repetition on achievement and school adjustment. Until the 2000s, a large body of published works in this area reported detrimental effects of grade repetition (Jackson 1975; Holmes 1984; Shepard and Smith 1989; Jimerson 2001, 2002). However, confidence in such findings has been shaken by mounting criticism, as most early studies failed to adequately control for pre-existing differences between retained and promoted students (Alexander et al. 2003; Lorence 2006). In the following, we review the recent literature regarding the substantive findings in earlier and later grades (Sections 2.1 and 2.2) and the empirical strategies employed for the identification of the causal effects (Section 2.3).

Since the practice of grade repetition differs across countries, its effects may vary across systems, and the findings from one context may not always be applicable to others. First, there is variation in timing. Grade repetition as a strategy to contrast poor performance is more common in primary school in some countries and in high school in others. Moreover, in some systems, it is used mainly at points of transition between educational cycles (e.g., between primary and middle school), with automatic promotion being the rule within cycles. Second, in some systems, grade repetition is determined by stringent rules based on standardized test scores, while in others school boards and teachers have ample leeway in deciding which students should be held back. Third, in high school, grade retention is much more common in tracked systems (where students are separated into schools providing different curricula) than in comprehensive systems where, in the main subjects, students are streamed by ability.[1]

Together, these studies suggest that grade repetition may be beneficial in primary school, but more harmful in later grades. Nonetheless, the evidence is still mixed and only based on data from a few countries.

2.1. Grade retention in early grades

Quite a few studies report beneficial effects of grade repetition in early grades. Three different articles (Jacob and Lefgren 2004, 2009; Allensworth 2005) analyzed the introduction of a test-based promotion policy in Chicago public schools. Using regression discontinuity design on administrative data, they found that grade repetition in third grade improved later achievement, while it had little effect in sixth grade. Greene and Winters (2007), and Schwerdt and West (2012) exploited a similar strategy in Florida, and both provided evidence of short-run beneficial effects of retention in third grade on reading and

[1] Although compulsory transfers to less demanding educational pathways are also possible in some tracked systems.

math scores. However, according to Schwerdt and West (2012), the advantage derived from repeating a grade faded out over subsequent years. For a cohort of low-achieving elementary students in Texas, Lorence et al. (2002) compared students held back in third grade to their promoted peers. The results suggested that grade repetition helped low-achieving students pass the state reading examination. In this case, an advantage persisted over the next six years. Examining the introduction of an automatic promotion policy in Brazilian primary schools, Koppensteiner (2013) identified an adverse effect on math achievement, arguing that the lack of the deterrent effect of grade repetition may have led to a decline in effort.

Several studies provide empirical evidence that grade repetition in the European context has positive effects. Fertig (2004) exploited data from a German retrospective survey to examine the academic outcomes of students held back in fifth grade, finding a positive effect of repeating a grade on educational outcomes. Similarly, Nunes et al. (2018), found a small positive impact of retention in fourth grade on sixth grade test scores in Portugal. Alet et al. (2013) estimated a short-run positive effect of grade repetition in third grade in France, but the advantage faded out or even reversed within four years.

Other studies report mixed results. Griffith et al. (2010) used the U.S. National Education Longitudinal Study and found that students retained between kindergarten and eighth grade showed lower progress in reading skills than their never retained peers. Two studies using data from Texas assessed the negative impact of grade repetition by matching a relatively small number of students' characteristics: Hughes et al. (2017) reported that grade repetition increased early school dropout and Wu et al. (2018) found that it was associated with lower progress in math skills. For students in New York City schools, Martorelli and Mariano (2018) exploited the policy of high stakes testing and found that grade repetition had no effect on behavioral outcomes in primary school. Pagani et al. (2001) found adverse effects on academic performance and behavioral adjustment in Quebec, reporting larger effects when grade repetition occurred at the early stages of primary school.

Turning to the European context, first-grade repeaters in Flemish schools outperformed equally at risk promoted peers the following year and were less likely to be held back throughout their remaining elementary school years (Goos et al. 2013). However, the achievement advantage disappeared in second grade as soon as repeaters faced new learning material. Eventually, the repeaters were less proficient than non-repeaters in math and reading. Grade repetition also had an adverse effect on psychological well-being, resulting in higher chances of transferring to special education in primary school. Bonvin et al. (2008) found contrasting outcomes in a matched sample of second-grade children in Swiss primary schools. Children repeating a grade performed better than their similar younger peers in the new class, but did not perform as well as comparable peers of their own age who had been promoted.[2] Focusing on self-concept, Kretschmann et al. (2019) examined a sample of students for three years after they were held back in sixth grade in Germany and found no evidence of positive effects. Another longitudinal survey in the Lisbon region carried out by Peixoto et al. (2016) showed that grade retention in sixth grade negatively affected both academic achievement and school emotional adjustment.

2.2. Grade retention in later grades

There are fewer studies on the effects of grade repetition in later grades. Jacob and Lefgren (2009) assessed the causal impact of grade retention on high school completion in the Chicago area. As already mentioned, they found that holding back younger students had no effect on their likelihood to finish high school; however, retaining low-achieving eighth grade students substantially increased their chances of dropping out of high school. Likewise, Eide and Showalter (2001) showed that repeating a year before tenth grade had adverse effects on dropout rates. Using data from the Chicago Longitudinal Study, Ou & Reynolds (2010) found that retention decreased the likelihood of enrolling in post-high school education. More specifically, late retention (between fourth and eighth grades) was more strongly linked to lower rates of post-secondary attainment than early retention (between first and third grades). In Louisiana, where retention is mostly determined by results on a high-stake test, Eren et al. (2017) observed that retention in eight grade increased the risk of dropout, and while analyzing the case of Columbia, Ferreira et al. (2018) reported that students retained in high school improved their reading but not their math test scores.

In the European context, Klapproth et al. (2016) compared the achievement of middle school repeaters in Luxembourg with that of their promoted peers. They found short-term benefits for those who were held back, although the two groups performed equally well in the medium run. However, grade repetition seemed to harm self-esteem. Ehmke et al. (2010) used longitudinal data for a sample of ninth graders in Germany and found no effect on math achievement or academic self-concept. Mathys et al. (2019) investigated the effect of grade repetition on a small sample of students in high school in French-speaking Belgium. They found decreased self-esteem, motivation, and perceived parental support in students held back in the seventh and eighth grades. Similar evidence is found in Cockx et al. (2018), who used data from a retrospective survey in Flanders, where underperforming students can choose either to repeat a grade or to enroll in a less demanding track; the downgrading option is a popular choice. They observed negative long-term impacts of grade repetition, including increased dropout rates,

---

[2] One debate that applies to some of the literature on the effect of grade retention is whether to compare students of the same age or in the same grade. The former compares retained students with promoted students of the same age who are attending the next grade. The latter compares retained students with promoted students who are one year younger but in the same grade. The results appear to vary systematically according to which approach is adopted (Goos et. al 2021). In this paper, we compare students of the same age. The reason is that we are analyzing the effect of retention after tracking, so it is unclear what would be a good match if the retained student changed tracks after being left behind.

schooling delays, and subsequent shifts to the vocational track. Another longitudinal survey from Flanders (Lamote et al. 2014) reported no effect on language achievement of retention in eighth grade, but a positive effect on academic self-concept. However, the effect of retention on language achievement became adverse in the longer run. Gary-Bobo et al. (2016) analyzed the effect of being retained in French middle school, finding that while grade retention has a small positive effect on test scores in ninth grade, it reduces the likelihood of reaching ninth grade in general (non-vocational) schools.

2.3. Empirical strategies

For ethical reasons, randomized experiments are not a feasible way to analyze the effects of grade retention (see Jackson 1975 for a few early exceptions). The main difficulty with non-experimental settings is that of deriving the counterfactual outcome of a student who has repeated a grade, if instead of being held back, he or she had been promoted. There is no one solution, because the appropriate strategy (if any) depends on the selection rules into treatment. Different methods have been applied in the literature, depending on the institutional context.

One setting for the study of grade repetition effects is test-based promotion. A transparent rule based on the result of a standardized test rules out discretion on the part of teachers and principals and allows estimation of a local treatment effect by means of regression discontinuity design techniques. Jacob and Lefgren (2004, 2009) exploited such a test-based promotion policy to assess the effect of grade repetition on students enrolled in the Chicago Public School System. Other who have employed this strategy include Schwerdt and West (2012), Eren et al. (2017) and Martorelli and Mariano (2018).

Another ideal situation occurs when a new policy comes into effect. For example, Koppensteiner (2013) exploits a change in the retention policy in primary school in Brazil to analyze the impact of the introduction of automatic promotion on school performance, while Ferreira et al. (2018) exploited timing differentials in the application of a change in the national retention policy in Columbia.

Identifying the effect of grade retention is more challenging in contexts where there is no test-based promotion policy and where policy reforms are not being made. Some scholars have employed instrumental-variable methods. For example, Alet et al. (2013) instrumented retention with the quarter of birth, while Fertig (2004) used physical development. The identifying assumptions are that time of birth and physical development may influence retention (older and more mature students are less likely to be retained) but have no direct effect on future outcomes given retention. Eide and Showalter (2001) have used exogenous variations in the starting age for kindergarten across U.S. states as instrumental variables. Other studies (Gary-Bobo et al. 2016, Cockx et al. 2018) use a broader framework and estimate dynamic models to estimate the effects of grade retention, controlling for unobservables and handling the initial conditions problem.

A number of scholars have used propensity score (PS) matching to assess the impact of grade retention on a variety of outcomes, such as future educational careers and psychosocial factors (Bonvin 2008; Wu et al. 2008; Ou and Reynolds 2010; Ehmke et al. 2010; Goos et al. 2013; Klapproth et al. 2016; Hughes et al. 2017; Mathys 2019). Propensity scores reduce potential bias by creating matched treatment and comparison groups with similar probabilities of being in the treatment group, conditional on a set of covariates. The strategy has been applied mainly in contexts where the promotion/retention decision is subject to some discretion, so that a credible counterfactual for retained students can be identified among existing promoted students. However, the validity of this strategy depends on data availability, in terms of sample size (a large N is needed to find good matches) and richness of information, as the matching model should include all possible confounders, both at the individual and at the school level, if contexts also play a role.[3]

3. The Italian case

The Italian educational system is organized into four cycles. Primary school starts at age 6 and lasts five years, after which pupils attend middle school from sixth to eighth grade. At age 14, students are tracked into different high schools, although they may also choose 2–3-year regional programs in the vocational education and training sector to fulfill the compulsory education requirement. High school lasts five years (from ninth to thirteenth grade) and offers several curricula, broadly categorizable into: traditional lyceums (academically oriented track), other lyceums (artistic, social science and linguistic), and technical and vocational high schools. Having a high school diploma qualifies students to enroll in tertiary education with no merit-based restrictions.

Despite the system’s openness, an extremely high level of social segregation characterizes high school choices in comparison Italy to other European countries (Jackson, 2013). Higher levels of parental education correlate strongly with successful early school performance, which increases a student’s likelihood to enroll in a lyceum. The direct effect of social background given performance is particularly large: given successful school performance, high-status families opt for lyceums much more frequently than low-status and migrant families (Contini and Scagni 2013; Panichella and Triventi 2014). Social stratification in high school choices is a critical issue, as the odds of entering and graduating from university are much lower for students with technical and vocational diplomas than for those who graduate from lyceums (Contini and Salza 2020).

The rules governing student progression differ at each level of education. Grade retention is extremely rare in primary schools, where students advance to the next grade almost automatically, even if they have not fully achieved the learning targets. Retention is slightly more common in middle schools, when teachers can decide to hold underachieving students back. At the end of the 2017-18 academic year, the

[3] The reported studies are based on small- to middle-sized samples, ranging from 500 to 1,500 students.

aggregate national rate of grade repetition was 2.1% in sixth grade, 1.7% in seventh grade, and 1.7% in eighth grade (ISTAT 2019).[4]

The situation changes drastically thereafter. Grade level retention is common among high school students, especially at the end of ninth grade; the percentage of students retained in high school is five to six times higher than in middle school, with large differences across tracks (Table 1).

Table 1. Retained Italian high school students by grade and track in academic year 2018-19

	9th	10th	11th	12th
High schools (Total)	11.8%	7.7%	7.7%	5.4%
Lyceums (Trad. and others)	7.9%	4.7%	4.8%	2.9%
Technical	16.7%	10.2%	10.4%	7.3%
Vocational (5- and 3-yr)	14.7%	12.1%	11.1%	9.2%

Source: Ministry of Education

In high school, evaluation of student performance is entrusted to the different subject teachers, according to 'the autonomous and competent expression of the teacher', and made in conjunction with the rest of the teaching staff. Students are assessed on their learning progress, behavioral conduct, and overall scholastic standing. At the end of the school year, the subject teachers assign final grades to their students. To be promoted, students have to get a minimum grade of 6 (out of 10) in each subject, including behavioral conduct. If they receive less than that in one or more subjects, retention is considered as an option for the next school year, depending on 'the chances the student will be able to meet the expected formative goals by the end of the year' (OM 92/2007). Students who are deemed unlikely to reach those goals are retained. Otherwise, an individual remedial learning plan is developed for the student, and the decision postponed to before the start of the following school year, after an *ad hoc* examination. The law makes explicit reference to teachers' expectations: as such, the retention decision is subject to a considerable degree of discretion and may vary substantially across schools.

4. Data

The empirical investigation in this paper relies on a novel longitudinal database that combines data from the National Register of Students (ANS) with that from the National Institute for the Evaluation of the Educational System (INVALSI). The construction of administrative archives collecting individual-level data of student careers is a rather recent practice in Italy. For this reason, when the data was released, it was considered reliable only from 2013 and only for certain regions. As a consequence, the data used in this study covers the entire population of students registered for high school during the school years

[4] When individual achievements are unsatisfactory in primary and middle school, schools must activate ad hoc recovery strategies. In "extraordinary cases," teachers may choose to retain a student if the decision is unanimous.

2013-14 through 2016-17 in the three largest and most populated regions in northern Italy (Lombardia, Piemonte and Veneto).[5]

We focus on students who entered high school (ninth grade) for the first time in the 2014-15 school year. This allows us to match students using data available from 2013-14, observe whether they were promoted or held back in 2014-15, and review their education records over the next two school years (2015-16 and 2016-17).

Sometimes, individuals disappear from the dataset. When this happens, we label them as a "dropout from the school system". Dropouts may include students who transferred from high school to regional vocational training programs (which are not included in the data archive). These students may also have moved abroad or to an Italian region not represented in the data. However, we consider this a negligible issue, as the share of students disappearing from the dataset is consistent with the regional statistics on early school leaving.

In addition to information about the students' grade progression and school attended, for each school year we track their end-of-year marks in behavioral conduct and in Italian and math, which are considered to be the two most important subjects. ANS also contains demographic information, such as gender, age, and migration status. INVALSI contributes other important details such as standardized test scores in reading and math literacy administered to students at the end of eighth grade, and information about parental socioeconomic background.

Table 2 shows descriptive statistics of the students enrolled in ninth grade in school year 2014-15 by retention status at the end of the year. To describe social background, we consider parental education (less than high school=neither parent has a high school degree; high school=at least one parent has a high school degree; higher education=at least one parent has a university degree). High schools are grouped into the four broad tracks already mentioned: traditional lyceums, other lyceums, technical high schools, and vocational high schools, ordered according to a widely accepted ranking from the most to the least academic-oriented.

Retained students disproportionately belong to families with low parental education and with migrant backgrounds, are more often boys and older. They are also overrepresented in technical and vocational schools, reflecting the fact that retention is more common in these tracks. Not surprisingly, promoted students outperform retained students in terms of both teachers' marks and standardized test scores. However, as shown by the large standard deviations, the performance distribution of promoted and retained students largely overlap, making it possible to successfully match students in the two groups.

---

[5] Data release and protection was regulated according to a protocol signed by the Italian Ministry of Education and the National Institute for the Evaluation of the Educational System (INVALSI).

Table 2. Descriptive statistics by retention status in ninth grade (school year 2014-15)

	Promoted students	Retained students	Total
Female	.52	.40	.51
Year of birth <2000 (older than expected age for the class)	.05	.20	.07
Parental education			
Less than high school	.29	.45	.32
High school	.44	.34	.42
Higher education	.20	.10	.19
Missing	.07	.11	.07
Migration background:			
Native	.91	.79	.90
2nd generation	.04	.08	.04
1st generation	.05	.13	.06
High school track			
Traditional lyceum	.41	.22	.39
Other lyceum	.13	.11	.12
Technical high school	.34	.48	.36
Vocational high school	.12	.19	.13
Performance measures in 8th grade:			
Teacher grade in math	7.46 (1.26)	6.28 (.62)	7.29 (1.26)
Teacher grade in Italian	7.48 (1.09)	6.41 (.65)	7.32 (1.11)
Teacher grade in behavior	8.92 (.85)	8.06 (.88)	8.79 (.91)
Math test score	.33 (1.02)	-.52 (.85)	.16 (.90)
Italian test score	.27 (.87)	-.51 (.76)	.21 (1.04)
Performance measures in 9th grade			
Teacher grade in math	6.68 (1.28)	4.15 (1.15)	6.36 (1.51)
Teacher grade in Italian	6.78 (.87)	5.2 (.96)	6.58 (1.04)
Teacher grade in behavior	8.76 (.84)	7.63 (.92)	8.61 (0.93)
N	108,369	17,867	126,245

NOTE. Standard deviations in parenthesis

5. Analytical strategy

5.1 Outcomes of interest

In this paper, we measure the impact of being held back in ninth grade on students' subsequent educational careers. We consider students who entered ninth grade for the first time in 2014-15 (year *t*). We analyze the impact of grade retention in year *t* on the educational path over the following two-year period. We focus on four key events: the enrollment decision at the beginning of year *t*+1, the end-of-year-result of year *t*+1, the enrollment decision at the beginning of year *t*+2 and the end-of-year-result of year *t*+2 (Figure 1). At the beginning of year *t*+1, students may either choose (outcome Y_1): to enroll in the same school[6], change institution while remaining in the same track, move to a less demanding track (in which cases the promoted students enroll in tenth grade and the retained students repeat ninth grade) or to leave the school system. During the school year, students may (outcome Y_2): drop out or continue their studies. In the latter case they could either be admitted to the next grade or fail. Similarly, we define two subsequent outcomes regarding the enrolment decision in school year 2016-17 (outcome Y_3) and the end-of-year result (outcome Y_4). Given the longitudinal nature of the observations, subsequent outcomes are defined only if individuals do not drop out of the sample first.[7,8]

Figure 1. Timeline of the key events

Middle school	**High school**				
Year *t*-1 **2013-14**	**Year *t*** **2014-15**	**Year *t* +1** **2015-16**		**Year *t* +2** **2016-17**	
8th grade	9th grade	9th grade if retained 10th grade if promoted		9th grade if retained twice 10th grade if retained once 11th grade if always promoted	
PERFORMANCE MEASURES Teachers' marks Test scores	TREATMENT DECISION Promoted N=108,369 Retained N=17,876	OUTCOME AT ENROLMENT No enrolment Down track change School change=track Same school	OUTCOME END OF YEAR Drop during year Prom. or Retained	OUTCOME AT ENROLMENT No enrolment Down track change School change=track Same school	OUTCOME END OF YEAR Drop during year Prom. or retained

We then analyze the overall path over the two-year time span. We classify the trajectories as follows: (i) *linear career,* when the student remains in the same school or in the same track and is not held back within the next two years; (ii) *downward change*, when he or she moves to a lower track and is not held back within the next two years; (iii) *retention*, when he or she experiences at least one grade repetition within the next two years; (iv) *dropout*, when at some point within the two-year observation window he or she is no longer present in the dataset.[9] It would have been particularly salient to evaluate the impact of grade retention on the probability of obtaining an upper secondary diploma within a given period of time, but unfortunately the dataset does not cover the entire period in question.

---

[6] In principle, some institutions deliver degree programs belonging to different tracks. To avoid confusion, when we say "same school" we mean that the student does not change institution or track.

[7] More specifically, Y_2 is defined only for students enrolling in any school in 2015/16, Y_3 only for those who have not dropped out during the previous scholastic year and Y_4 for those enrolling in any school in year 2016/17.

[8] We exclude from the analysis students who drop out but then re-enrol in year *t*+1.

[9] A student who was held back and later dropped out is classified as dropout.

5.2 Identification of the causal effect of retention

The estimation of causal effects requires a comparison of potential outcomes under two or more distinct treatment conditions. In a one-treatment framework, the causal effect for a specific unit is the difference between the outcome this unit experiences by being exposed to treatment and what that outcome would have been without treatment (Imbens and Rubin 2015). In this study, we aim at estimating the average difference between the retained student's actual educational outcome Y_T and the counterfactual Y_C, representing what the career would have been like if he or she had been promoted. More formally, we can refer to the average treatment effect for the treated as:

$$ATT = E[Y_T - Y_C | T = 1]$$

where T is the binary indicator for exposition to grade retention. Since the counterfactual Y_C cannot be directly observed, it must be appropriately estimated.

In general, promoted and retained students have different characteristics, particularly in terms of school performance. In an 'ideal' environment where retention is assigned according to school performance assessment based on transparent and non-discretional rules, we could adopt a regression discontinuity design and identify a local causal effect of retention by comparing students just below and just above the performance threshold.

In principle, a well-defined treatment assignment rule does exist in our setting and is based on school grades. Students failing one or more subjects are not admitted to the next year. However, the decision is subject to considerable discretion: teachers and principals may hold different beliefs about the usefulness of retention and set school-specific standards. In practice, marks – integer numbers between 0 and 10 – are adjusted on the basis of the retention/promotion decision: if teachers wish to promote a student who has not fully satisfied the requirements, they can (slightly) adjust the final mark by assigning a 6 (pass) instead of a 5 (fail).[10]

This has an advantage and a disadvantage. The advantage is that this margin of discretion means that there are promoted students and retained students who have the very same ability, which provides theoretical grounds for using matching strategies. The disadvantage is that, due to endogeneity of current performance, we cannot match students on achievement measures that portray students just before the promotion/retention decision; we have to rely on prior performance records instead.

To match students, we exploit a large array of performance measures taken at the end of eighth grade. These include school marks in Italian and math – incorporating students' curricular knowledge, competences, and effort – and scores on tests designed to specifically measure competences, following a similar conceptual framework as that set out in PISA (OECD 2019). Also available is the teacher-

[10] In principle, marks vary between 0 and 10, with 6 being the minimum needed to pass. In practice, however, marks below 4 are hardly ever given on the final report card.

assigned mark in conduct, which reflects students' diligence in carrying out school tasks and their behavior, thus capturing non-cognitive traits such as self-control and motivation.

Let us formalize the problem. In principle, treatment assignment depends on the following deterministic model:

$$T_{ij}^* = \beta_0 + \beta_1 A_{ij} + \beta_2 X_{ij} + \beta_3 Z_j + \mu_j \quad (1)$$
$$T_{ij} = I(T_{ij}^* > 0)$$

where T^* is a retention score according to which the decision is taken and T is the binary treatment variable, Z captures school-level observed characteristics potentially affecting school learning standards (school curriculum and student body composition in terms of prior ability and family background). μ is a latent index of the school's propensity to use retention as a remedial measure, assumed orthogonal to the other explanatory variables. We refer to a school's propensity to retain, conditional on the school observable characteristics, as *strictness*. Within each school, the decision to promote or hold students back depends primarily on the students' current ability vector A, which, given the richness of the observed measures, captures a wide range of cognitive and non-cognitive abilities. However, other sociodemographic characteristics X may also play a role: as shown in Salza 2022, even net of prior abilities, students from disadvantaged backgrounds are more likely to be retained. As motivated above we cannot include contemporary school grades in the model because they are endogenous to the retention decision; thus, we rely on the closest available prior achievement measures (school marks and test scores one year before, at the end of eighth grade).

We may rewrite expression (1) as:

$$T_{ij}^* = \beta_0 + \beta_1 (A_{-1ij} + \Delta A_{ij}) + \beta_2 X_{ij} + \beta_3 Z_j + \mu_j \quad (2)$$
$$= \beta_0 + \beta_1 A_{-1ij} + \beta_2 X_{ij} + \beta_3 Z_j + \mu_j + \varepsilon_{ij}$$

where A_{-1} is prior ability and ΔA is the progress made in the current year. Since ΔA is unobserved to the researchers, $\varepsilon_{ij} = \beta_1 \Delta A_{ij}$ is an individual unobservable component.

Similarly, for explanatory purposes we may define a stylized outcome model as:

$$Y_{ij}^* = \gamma_0 + \gamma_1 A_{ij} + \gamma_2 X_{ij} + \gamma_3 Z_j + \theta T_{ij} + u'_{ij}$$
$$= \gamma_0 + \gamma_1 A_{-1ij} + \gamma_2 X_{ij} + \gamma_3 Z_j + \theta T_{ij} + u_{ij} \quad (3)$$

assuming that the outcome may depend on treatment and on the same individual characteristics affecting the retention decision.[11] The error term $u_{ij} = \gamma_1 \Delta A_{ij} + u'_{ij}$ includes the new ability contribution and an idiosyncratic error term. To simplify the exposition, in (3) we assume that retention has the same

[11] Differently from few contributions exploiting IV strategies to analyse the effects of retention, we cannot think of individual characteristics affecting retention but not directly affecting subsequent performance.

impact θ on all students; however, in additional analyses we will allow for heterogeneous effects across school tracks and family background. Moreover, this specification might be an oversimplification of the true outcome model, as various interaction effects and non-linearities might be in place.[12]

To estimate the causal effect of grade repetition we use a matching strategy. Ideally, the aim would be to find a control unit for each treated unit who has the same performance measures and social and migratory background and who attends a similar type of school (same track and same social and ability composition). However, it is impossible to find a perfect match for all the variables of interest in real contexts. PS matching allows matching to be made according to a scalar value, representing the predicted treatment probability given observed individual characteristics. PS matching delivers unbiased estimates of the treatment effect if the treatment assignment model includes all the relevant confounding variables between the treatment and the outcome, so that, given the propensity score, treatment and outcome are orthogonal (CIA-Conditional Independence Assumption). In other words, this assumption implies that we may assume random assignment of the treatment within subpopulations defined by values of the covariates. The second key assumption is that for large samples we can find both treated and control units for all values of the covariates (Common Support). In practice, this means that we must be able to find sizable groups of treated and control units with very similar propensity scores.

However, CIA is challenged if we use prior performance measures rather than current performance. Since the larger the ability progress ΔA_{ij}, the lower the retention probability, when neglecting progress, we tend to overestimate the negative (or underestimate the positive) impact of treatment. We will come back to this issue below. Note that although most other studies use previous performance measures to control for ability, to our knowledge, ours is the first study to make explicit the issues involved.

Another question is related to the unobserved school strictness μ. In this multilevel framework – individuals are nested into schools – the most effective strategy for maximizing the probability of finding good matches in terms of both individual and context characteristics would be to search for the match within the same school, or in schools with the same observed characteristics and degree of strictness. If the sample size is too small, a good strategy is to rely on PS estimation based on cluster (here schools) fixed effects models (Arpino and Mealli 2011).

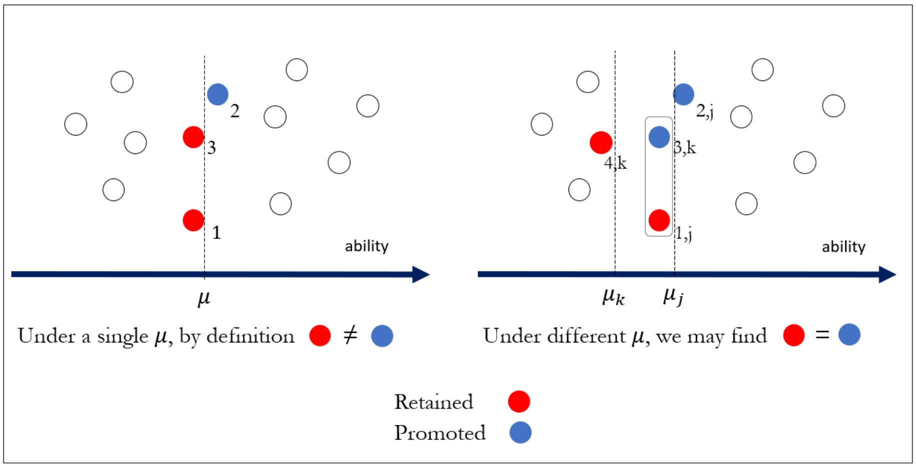
Nevertheless, if learning standards and strictness are set at the school level, there are theoretical reasons *not* to choose matching units within the same school. In principle, if a student with given characteristics and ability is retained, since decision rule (1) is ideally deterministic, we would not be able to find a

[12] We also assume that the unobserved school strictness component does not affect future outcomes, conditional on other school characteristics and the actual retention decision. In Table A5.1 in the Appendix, we display the results of the estimation of model (3) on all matched units, with the addition of estimated strictness: the evidence is that strictness is not related to the probability of linear, downward changes and dropout outcomes and only slightly positively related to the probability of experiencing (another) retention.

promoted student with the same characteristics and abilities in the very same school. If we did find a good match on the observable component of ability, these students would necessarily differ in the unobservable component ΔA_{ij}, and the promoted student would have more favorable characteristics than the retained one.

Against this background, we exploit variations in school strictness: the possibility of finding an appropriate match rests on the search for each retained student of a promoted student displaying the same observed performance among ideally identical schools – in terms of curricula, social and ability composition – but with a lower propensity to retain (see Fig.2). School strictness μ_j can be estimated from model (2), as posterior Bayesian estimates within a multilevel logistic model with random intercept at the school level. To ensure consistent estimation of school strictness, we assume that initial high school choices are based on observed school characteristics (school type, student body composition in terms of social background and abilities), but not on the unobserved orthogonal component of strictness itself.[13] In the light of the large array of available ability measures (school marks, test scores and behavior), we believe this is a credible assumption.

Figure 2. Finding a comparable match: same school or different school



Left panel: With the same level of strictness, it is impossible to find students with the same latent ability *A* in the treated and control groups. Units 1 and 3 perform below the threshold and are retained; unit 2 performs above the threshold and is promoted. **Right panel**: With differences in strictness, it is possible. Although units 1 and 3 have the same ability, unit 1 is below the threshold in school *j* and therefore retained; instead, 3 is above the threshold in school k and therefore promoted.

An interesting consequence of this strategy is that by imposing a match with control pupils in less strict schools, we also tackle the confounding problem caused by the (inevitable) omission of current performance. If grade retention worsens future outcomes ($\theta < 0$), as will result in the empirical

[13] If initial school choices depend on unobserved strictness there might be a positive correlation between μ and ε, yielding to biased estimates of μ.

analyses, by forcing the comparison of treated units in stricter schools with control units in less strict schools we end up attenuating the unbalance existing in the "new ability" component and potentially even reversing it (see proof in Appendix 4).[14]

5.3 Matching procedure

Our matching strategy combines exact matching and PS matching and limits the search to the school displaying the least degree of strictness, within an acceptable range of the predicted retention probability.

The following steps are taken:

1) We exploit the large *N* in our data to find for each retained student a perfect match among promoted students on: (i) the individual variables of gender, age, parental education, migration background; (ii) high school track and the school percentile in math according to test scores in eighth grade, to control for the school composition in terms of ability.

2) Using the data from the entire student cohort, we estimate model (2), with a two-level logistic model with random intercepts at the school level, including all the individual and school-level explanatory variables detailed in Table 3. In addition, we compute Bayesian point estimates of school strictness.

3) For all retained pupils finding an exact match on the corresponding variables, the identification of the pool of comparable controls works as follows. First, we evaluate the probability of being retained for each student, given her individual and school characteristics at a given degree of strictness, say $\mu_j = 0$, by computing:

$$\hat{P}\,(T_{ij} = 1|\mu_j = 0) = 1/\left[1 + exp\left(\hat{\beta}_0 + \hat{\beta}_1 a_{1ij} + \hat{\beta}_2 x_{1ij} + \hat{\beta}_3 z_j\right)\right] \qquad (4)$$

Second, for each retained student we identify the pool of potential controls. Following the guidelines in the existing literature on PS matching (Stuart and Rubin 2008), we include all promoted students who are in the range of 0.20 standard deviations from the p-score of the retained. Third, within this set of potential controls, allowing for multiple matching of control students with different treated units, we choose as the best match the promoted student enrolled in the school displaying the *least degree of strictness*, provided that the school strictness is lower than that of the corresponding retained student. This means that if there are no promoted students within the desired PS range in less strict schools, the treated unit will not be matched. Fourth, using only matched units, we compute the share of students exhibiting each of the conditional outcomes (Figure 1) and the unconditional outcomes described above.

As sensitivity checks, we set even more restrictive rules to search for the match: (i) we allow for a smaller range (0.10 s.d.) for the propensity scores; (ii) within both ranges - 0.20 and 0.10 s.d. - we

[14] If grade retention instead improved future outcomes, the strategy would either deliver less biased underestimates of the positive effect of treatment, or overestimates of the effect.

impose unilateral intervals where the individual retention probability is also forced to be higher for the controls than for the treated students. In this latter case, we end up matching retained students attending stricter schools *and* with a lower retention probability given strictness than the homologous promoted students. For comparison, we also estimate the impact of retention with similar-strictness matching. After ordering schools from the least to the strictest, we restrict the search for a match in the least strict school within the closest 5th-percentile of the strictness range. While with similar-strictness matching retained and promoted students will display very similar average grades and test scores, we expect that matching treated units with control units in less strict schools will unbalance observed ability measures in favour of the treated, allowing to "correct" for possible differences in unobserved ability that could lead to an overestimation (underestimation) of the negative (positive) effects of retention (see Table A6.1 in the Appendix)

Table 3. Description of the variables

INDIVIDUAL VARIABLES		
Age	1=whether the student is older than the modal age of the classroom; 0=otherwise	Exact matching
Parental education	Highest level of education attained by the mother and/or father: up to middle school, high school diploma, tertiary education	Exact matching
Migration background	Native citizens: born in Italy, Italian citizenship Second-generation migrants: born in Italy, foreign citizenship First-generation migrants: foreign-born, foreign citizenship	Exact matching
Gender	1=female, 0=male	Exact matching
Performance	Final marks in Italian, math and conduct (behavior), 8^{th} grade Test scores in Italian and math, 8^{th} grade	Propensity score
HIGH SCHOOL VARIABLES		
Track	Traditional lyceums, other lyceums, technical high schools, vocational high schools	Exact matching
Ability composition: school percentile in math test scores	According to math test scores measured at the end of 8^{th} grade	Exact matching
Ability composition: average teacher marks and INVALSI scores + standard deviation in math INVALSI score	According to final teacher marks and INVALSI scores in 8^{th} grade	Propensity score
Socio-demographic composition	School percentages of females, migrants, students older than the median age in 9^{th} grade, and students with highly educated parents (tertiary degree)	Propensity score

6. Results

A total of 179,350 students were enrolled in ninth grade in the school year 2014-15 in the regions covered by the data. We exclude those who were not attending ninth grade for the first time (N=23,285), as well as those with missing data on the outcomes of interest or with missing performance data. We end up with a final analytical sample of 126,245 students, of whom 17,876 were retained (Table A2.1 in the Appendix).

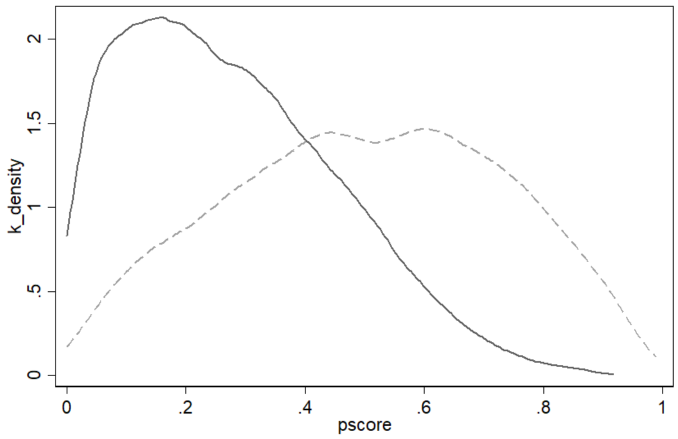
Following the steps described in Section 5.3, we successfully match 7,453 of the 17,876 (41.91%) retained students in the analytical sample to comparable promoted students (4,106 distinct individuals, as 3,347 promoted students are matched with more than one retained student). We exclude from analysis students who disappear from the panel and are then observed again (N=59). This leaves us with a total of 7,438 retained students matched with 7,410 promoted students.

Thus, less than half of the students held back found a close match. In other words, many retained students performed far too poorly to find a credible match among promoted students. This is shown in Figure 3, where we compare the PS distribution, computed from (3), of matched and non-matched retained students: clearly, matched students tend to have a much lower PS, and this is largely driven by differences in performance.

Hereinafter, to qualify our estimates we adopt the term "LATT", meaning the local average treatment effect of the treated. What we are estimating in practice in fact is a *local effect* because what we are able to estimate is the effect on the treated, near the (fuzzy) threshold between the pass or fail decision.

In Figure A3.1 in the Appendix we show the PS distributions for promoted and retained students in the whole student population (left panel) versus the matched sample (right panel); in the latter we observe a very good balance between treated and controls, as the corresponding PS distributions basically overlap.

Figure 3. PS distributions within retained (matched vs. non-matched)



Solid line: matched retained students
Dashed line: non-matched retained students

In line with our expectations, forcing retained students to be matched to promoted students in less strict schools, results in a treatment group composed of slightly better performing individuals than in the control group (Table 4). Retained students in the matched subsample outperform the promoted ones on all five measures of performance: they display higher average teacher marks in Italian, math, and behavior and higher test scores (closer to the population mean, equal to 0) in the reading and math literacy standardized national assessments. In this way, we contrast the potential bias related to the impossibility to control for the difference in progress made during ninth grade between the treatment groups described in Section 5.2. *Ceteris paribus*, this strategy should lead toward conservative estimates of the effect of grade repetition if the effect turns out to be detrimental, or to an overestimate of the effect if it turns out to be beneficial.

Table 4. Teacher marks and standardized test scores in treatment groups (matched subsample)

	Treated	*Control*	*Standardized bias*
Italian mark	6.62 (0.76)	6.46 (0.74)	0.22
Math mark	6.45 (0.76)	6.35 (0.75)	0.13
Behavior mark	8.32 (0.80)	8.06 (0.91)	0.33
Reading test	-0.25 (0.81)	-0.45 (0.84)	0.28
Math test	-0.27 (0.71)	-0.43 (0.76)	0.23
N	7438	4106	

NOTES. Standard deviations in parenthesis. The standardized bias is the difference in the mean value of a variable between the treatment and control groups, divided by the standard deviation in the treatment group.

We now turn to the results. Table 5 displays the estimated LATT expressed as probability differences and relative risks (risk ratios). We report the percent experiencing each event of interest at the beginning and at the end of the two school years after year *t* (i.e., ninth grade in 2014-15), and hence the outcome probabilities are conditional on survival (i.e., not having dropped out) up to that point.

Retention strongly affects the careers of students. 9.7% of the retained students do not enroll in school at the beginning of year t+1 as opposed to 0.4% of the students in the control group, so the non-enrollment relative risk is 25.6. Compared to non-retained students, repeaters are 4.4 times more likely to transfer to a different school (14.3% vs. 3.3%), and 11.6 times more likely to switch to a lower track (22.3% vs. 1.9%). Only about half of the repeaters keep attending the same school as opposed to 94% of comparable students who were promoted. By the end of the year, roughly 8.2% of the enrolled repeaters had left high school (5 times more likely than non-repeaters). On the other hand, the share of students failing year *t*+1 is slightly lower for repeaters. This is not surprising, as teachers might be reluctant to hold students back twice in a row. When looking at the second year (*t*+2), the gap with comparable promoted students narrows somewhat. However, repeaters are still six times more likely not to enroll than matched promoted students. Moreover, 5.6% of repeaters move to a lower track, as opposed to 2.6% of controls; even more importantly, 6.6% of the surviving treated students drop out from high school at the end-of-year *t*+2, vs. 2.8% of the similar controls. At this point, the probability for grade repetition is also slightly higher for retained students than for the controls (14.4% vs. 12.6%).

Table 5. LATT estimates. Conditional outcomes

	Treated		Control				
Year t+1	N	%	N	%	Prob Diff	Prob Ratio	P-value
At enrollment	7,438		7,410				
Non-enrollment (dropout)		9.7		0.4	9.3	25.6	0.000
Track (downward) change		22.3		1.9	20.4	11.6	0.000
School change (same track)		14.3		3.3	11	4.4	0.000
Same school		53.7		93.9	-40.2	0.6	0.000
End-of-year outcome	6,719		7,382				
Dropout during the year		8.2		1.6	6.6	5.0	0.000
Grade repetition		14.6		15.9	-1.3	0.9	0.040
Promoted		77.2		82.5	-5.3	0.9	0.000
Year t+2							
At enrollment	6,169		7,262				
Non-enrollment (dropout)		8.2		1.4	6.8	6.0	0.000
Track (downward) change		5.6		2.6	3	2.1	0.000
School change (same track)		4.8		5.4	-0.6	0.9	0.000
Same school		81.3		90.6	-9.3	0.9	0.000
End-of-year outcome	5,662		7,162				
Dropout during the year		6.6		2.8	3.8	2.3	0.000
Grade repetition		14.4		12.6	1.8	1.1	0.002
Promoted		79.0		84.6	-5.6	0.9	0.000

Table 6 provides an overview of the two groups' school careers after school year 2014-15. The outcome is now characterized as the entire school path in the two school years following the retention decision in ninth grade. We classify the paths as follows: (i) a *linear career*, when the student never changes school and is never retained; (ii) a *downward change*, when the student moves to a lower track and is thereafter promoted; (iii) *retention*, when the student is held back; (iv) *dropout*, when at some point within the two-year observation window the student is no longer observed in the panel.[15]

The share of students experiencing a linear career is much lower for the treated (37.2%) than for the controls (69.5%). Retained students are more likely to make a downward change (+18.4 percentage points, i.e., about 9.2 times the comparison group) and to drop out from high school (+21.8 percentage

[15] If a student experiences grade retention and drops out afterwards, he or she is classified as a dropout.

points, i.e., 6 times the comparison group). Grade repetition is less common among the treated (-7.8 percentage points, risk ratio 0.7).

Table 6. LATT estimates. Unconditional outcomes

	Treated (%)	Control (%)	Prob Diff.	Prob. Ratio	p-value
Linear	37.2	69.5	-32.4	0.5	0.000
Downward	20.6	2.2	18.4	9.2	0.000
Retention	16.0	23.9	-7.8	0.7	0.000
Dropout	26.2	4.3	21.8	6.0	0.000

6.1 Heterogeneity of the effects

We now examine the treatment effect by high school track in ninth grade, parental education and migration background. We do so by looking at differences in the probability of experiencing each of the four trajectories described above, for treated and controls in different social groups. The statistical significance of these interaction effects is shown in the Appendix.

For students enrolled in traditional lyceums, high parental education significantly mitigates the negative effects of retention (Figure 4 and Table A7.1 in the Appendix). More specifically, the probability of a linear career (top-left panel) is substantially higher for retained pupils whose parents have tertiary education than for those with lower education, and the difference between the corresponding probabilities for treated and controls is smaller. Similarly, for the most advantaged group we observe a lower effect of grade retention on the probability of experiencing a downward career (top-right panel) and of dropping out (bottom-right panel), although the difference is not statistically significant. There is also some evidence of a slightly lower impact of grade retention for students with better educated parents in technical schools. Instead, for students in non-traditional lyceums and vocational schools, the negative effects of grade retention do not appear to differ across parental education levels.

Figure 5 analyses grade retention effects by migration background. There is some evidence that the grade retention effect in vocational schools in terms of the probability of not experiencing a linear career is higher for migrant students than natives (Table A7.2 in the Appendix).

6.2 Sensitivity checks.

The results of the sensitivity checks described in Section 5.3 are reported in the Appendix (Table A6.3). We focus on the impact estimates on career paths (unconditional outcomes). Overall, our findings are consistent across matching procedures. Our preferred estimates based on bilateral propensity score and the choice of the least strict school within that range (provided that the school of the control unit displays a lower degree of strictness of the school of the treated unit) reveal that the risk of dropping out from the school system is six times higher for retained students than for matched promoted students. The estimates vary between 5.5 and 6.1, depending on the procedure. The corresponding estimate with similar-strictness matching, displayed for comparison purposes, would be 7.9. Thus, in line with expectations, less strict matching delivers more conservative estimates of the (negative) effect of retention, as compared to similarly-strictness matching.

Figure 4. Shares of matched treated and controls experiencing different 2-year paths, by school track and parental education

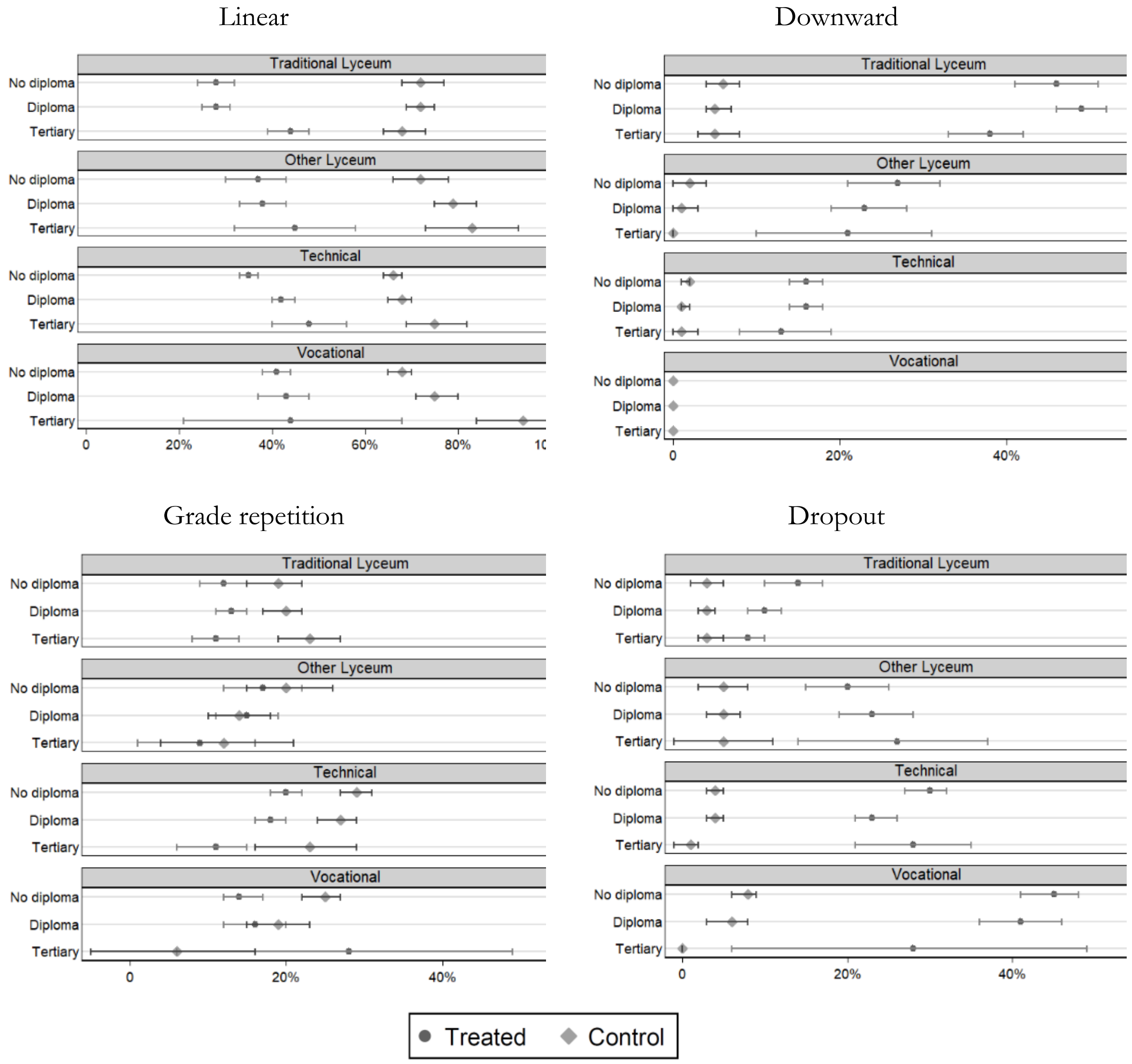
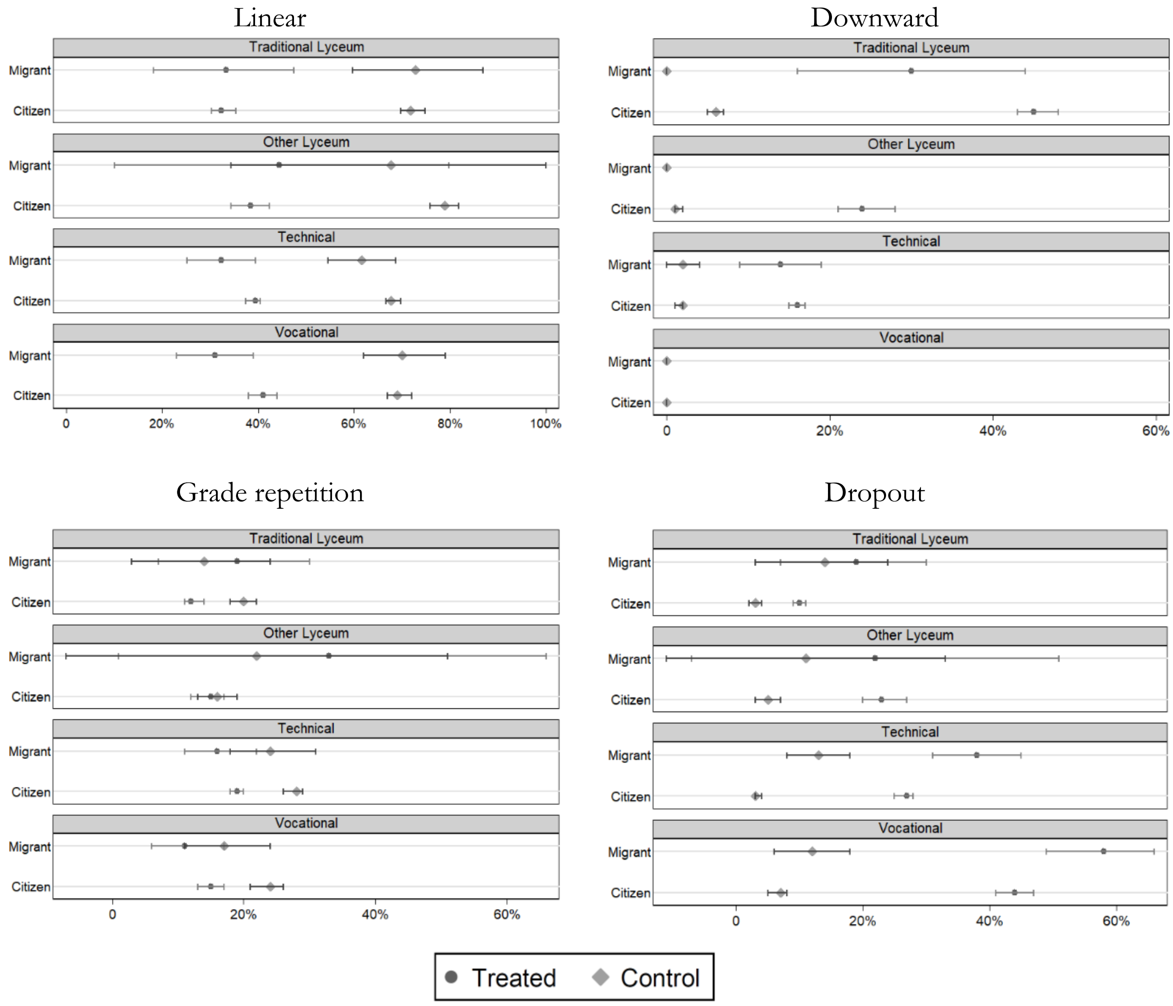


Figure 5. Shares of matched treated and controls experiencing different 2-year paths, by school track and migration background



7. Summary and conclusions

The practice of grade retention continues to be a contentious issue. Despite its widespread adoption in Italian schools, research into its effects has been limited by a lack of adequate data. This is the first study to undertake systematic analysis of the impact of grade retention in Italian high schools. We use a novel database that combines data from two administrative sources on students' educational paths and covers the entire population of secondary school students between school years 2013-14 and 2016-17 in the three largest regions in northern Italy. The database (N=126,245) includes a large array of performance measures, capturing competencies, disciplinary knowledge, effort, and diligence prior to the occurrence of grade repetition, which allows meaningful like-with-like comparisons, and afterwards, which allows analysis of the consequences on educational outcomes. Since individual data is clustered by school, we are also able to account for school-level factors in terms of social and ability composition that may affect both the probability and the impact of retention.

We propose a matching strategy to assess the causal impact of grade repetition in Italy, where there is substantial leeway in whether to promote or retain students. We argue that it is not theoretically possible to find fully comparable retained and promoted students within the same school. However, given that schools differ in terms of "strictness" – defined as the propensity to hold students back, given observable characteristics such as the school curriculum (track), the student body ability and socio-demographic composition – it is possible to find fully comparable retained and non-retained students. To ensure a good match, for each retained unit we look for a promoted student with the same array of ability measures (and other individual and school characteristics) in schools displaying a *lower* degree of strictness. We end up matching 42% of the entire population of retained students. Naturally, not all the retained students can be matched, because there is no corresponding control unit for those whose performance is too weak. In this perspective, we can interpret our results as estimates of a local average treatment effect on the treated (LATT).

Our results do not speak in favor of grade repetition. In the two years following ninth grade, we find that retained students face a much higher dropout risk than their promoted peers. Repeaters are also more likely to move to tracks with less demanding learning targets. Although less grave than dropping out, downward track transfers may result in loss of human capital, since earning a technical or vocational school diploma strongly reduces the chances of entering and completing tertiary education.

Moreover, there is evidence that the negative impact of grade retention is larger for the least advantaged social groups. Since students from families with low parental education and migrant backgrounds are more vulnerable to the risk of grade retention even net of prior performance (Salza 2022), we may conclude that – in addition to being ineffective as a remedial tool for poor performance – the practice of grade retention also exacerbates educational inequalities.

A note on external validity is in order. This study is based on population data from three large northern Italian regions. If not only institutional characteristics matter but also the environment in which students live, then the gap between northern and southern Italy in terms of economic development means that our estimates may not apply to the entire country. Southern Italy is characterized by lower levels of achievement (Argentin et al 2017), higher rates of early school leaving, youth unemployment, and NEET (Openpolis 2019), while retention rates are only slightly higher (MIUR 2019). We can expect grade retention to have a major negative impact even in the most disadvantaged areas of the country. More research is needed to determine how grade retention affects the lives of youths in different settings within a specific institutional context.

References

Alexander, Karl L., Doris R. Entwisle, and Susan L. Dauber. n.d. “Grade Retention: Lingering Questions.” In *On the Success of Failure. A Reassesment of the Effects of Retention in the Primary Grades.* Accessed January 21, 2017.

Alet, Élodie, Liliane Bonnal, and Pascal Favard. 2013. “Repetition: Medicine for a Short-Run Remission.” *Annals of Economics and Statistics*, no. 111/112: 227.

Allensworth, Elaine M. 2005. “Dropout Rates After High-Stakes Testing in Elementary School: A Study of the Contradictory Effects of Chicago’s Efforts to End Social Promotion.” *Educational Evaluation and Policy Analysis* 27 (4): 341–64.

Argentin Gianluca, Gianna Barbieri, Patrizia Falzetti, Emmanuele Pavolini, Roberto Ricci, 2017. “I divari territoriali nelle competenze degli studenti italiani: tra fattori di contesto e ruolo delle istituzioni scolastiche*.” Il Mulino-Riviste Web*, Fascicolo 1

Arpino, Bruno, and Fabrizia Mealli. "The specification of the propensity score in multilevel observational studies." *Computational Statistics & Data Analysis* 55.4 (2011): 1770-1780.

Belot, Michèle, and Vincent Vandenberghe. 2014. “Evaluating the ‘Threat’ Effects of Grade Repetition: Exploiting the 2001 Reform by the French-Speaking Community of Belgium.” *Education Economics* 22 (1): 73–89.

Bonvin, P., G. Bless, and M. Schuepbach. 2008. “Grade Retention: Decision-Making and Effects on Learning as Well as Social and Emotional Development.” *School Effectiveness and School Improvement* 19 (1): 1–19.

Bropy, Jere, International Institute for Educational Planning, and International Academy of Education. 2006. *Grade Repetition*. Paris; Brussels: International Institute for Educational Planning; International Academy of Education.

Checchi, Daniele, Carlo Fiorio and Marco Leonardi (2013), “Intergenerational persistence of educational attainment in Italy”, *Economic Letters*, vol 188, issue 1, 229-232

Cockx, Bart, Matteo Picchio, Stijn Baert (2018). “Modeling the Effects of Grade Retention in High School.” *Journal of Applied Econometrics*

Contini, Dalit, Guido Salza. 2020. “Too Few University Graduates. Inclusiveness and Effectiveness of the Italian Higher Education System.” *Socio-Economic Planning Sciences* 71 (September): 100803.

Contini Dalit, Andrea Scagni 2013 Social Origin Inequalities in Educational Careers in Italy. Performance or Decision Effects? in Jackson Michelle (eds), *Determined to Fail? Performance, Choice and Education*, Stanford University Press

Cosnefroy, Olivier, Rocher, Thierry. 2005. *Le redoublement au cours de la scolarité obligatoire: nouvelles analyses, mêmes constats*. Paris : Ministère de l’éducation nationale, de l’enseignement supérieur et de la recherche, Direction de l’évaluation et de la prospective.

Crahay, Marcel, and Géry, Marcoux. 2008. “Mais Pourquoi Continuent-Ils à Faire Redoubler? Essai de Compréhension Du Jugement Des Enseignants Concernant Le Redoublement.” *Revue Suisse Des Sciences de l’éducation* 30 (3): 501–18.

Eide, Eric R., and Mark H. Showalter. 2001. “The Effect of Grade Retention on Educational and Labor Market Outcomes.” *Economics of Education Review* 20 (6): 563–76.

Fertig, Michael. 2004. “Shot across the Bow, Stigma or Selection? The Effect of Repeating a Class on Educational Attainment.” IZA discussion paper 1266

Goos, Mieke, Jan Van Damme, Patrick Onghena, Katja Petry, and Jerissa de Bilde. 2013. “First-Grade Retention in the Flemish Educational Context: Effects on Children’s Academic Growth, Psychosocial Growth, and School Career throughout Primary Education.” *Journal of School Psychology* 51 (3): 323–47.

Griffith, Catherine Aurentz, John Wills Lloyd, Kathleen Lynne Lane, and Melody Tankersley. 2010. “Grade Retention of Students During Grades K–8 Predicts Reading Achievement and Progress During Secondary Schooling.” *Reading & Writing Quarterly* 26 (1): 51–66.

Holmes, C Thomas, and Kenneth M Matthews. 1984 “The Effects of Nonpromotion on Elementary and Junior High, Pupils: A Meta-Analysis”. *Review of Educational Research*. 1984;54(2):225-236.

Hughes, Jan N., Qian Cao, Stephen G. West, Paula Allee Smith, and Carissa Cerda. 2017. "Effect of Retention in Elementary Grades on Dropping out of School Early." *Journal of School Psychology* 65 (December): 11–27.

Imbens, W. Guido, Donald B. Rubin 2015, *Causal Inference in Statistics, Social, and Biomedical Sciences*, Cambridge University Press

Jackson, Gregg B. 1975. "The Research Evidence on the Effects of Grade Retention." *Review of Educational Research* 45 (4): 613.

Jackson, Michelle, ed. *Determined to succeed? Performance versus choice in educational attainment*. Stanford University Press, 2013.

Jacob, Brian A., and Lars Lefgren. 2002. "Remedial Education and Student Achievement: A Regression-Discontinuity Analysis." NBER Working Paper.

———. 2009. "The Effect of Grade Retention on High School Completion." *American Economic Journal: Applied Economics* 1 (3): 33–58.

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

Jimerson, Shane R., Gabrielle E. Anderson, and Angela D. Whipple. 2002. "Winning the Battle and Losing the War: Examining the Relation between Grade Retention and Dropping out of High School." *Psychology in the Schools* 39 (4): 441–57.

Klapproth, Florian, Paule Schaltz, Martin Brunner, Ulrich Keller, Antoine Fischbach, Sonja Ugen, and Romain Martin. 2016. "Short-Term and Medium-Term Effects of Grade Retention in Secondary School on Academic Achievement and Psychosocial Outcome Variables." *Learning and Individual Differences* 50 (August): 182–94.

Koppensteiner, F. Martin. 2014. "Automatic Grade Promotion and Student Performance: Evidence from Brazil." *Journal of Development Economics* 107 (March): 277–90.

Lorence, Jon, A. Gary Dworkin, Laurence A. Toenjes, Antwanette N. Hill, Andrew Rotherham, and Lorrie A. Shepard. 2002. "Grade Retention and Social Promotion in Texas, 1994-99: Academic Achievement among Elementary School Students." *Brookings Papers on Education Policy*, no. 5: 13–67.

Lorence, Jon. 2006. "Retention and Academic Achievement Research Revisited from a United States Perspective." *International Education Journal* 7 (5): 731–77.

Manacorda, Marco. 2008. *The Cost of Grade Retention*. London: Centre for Economic Performance, London School of Economics and Political Science.

Mathys, Cécile, Marie-Hélène Véronneau, and Aurélie Lecocq. 2019. "Grade Retention at the Transition to Secondary School: Using Propensity Score Matching to Identify Consequences on Psychosocial Adjustment." *The Journal of Early Adolescence* 39 (1): 97–133.

OECD (2019), PISA 2018 Results, PISA, OECD Publishing, Paris.

Pagani, Linda, Richard E. Tremblay, Frank Vitaro, Bernard Boulerice, and Pierre Mcduff. 2001. "Effects of Grade Retention on Academic Performance and Behavioral Development." *Development and Psychopathology* 13 (2): 297–315.

Panichella, Nazareno, and Moris Triventi. "Social inequalities in the choice of secondary school: Long-term trends during educational expansion and reforms in Italy." *European Societies* 16.5 (2014): 666-693.

Peixoto, Francisco, Vera Monteiro, Lourdes Mata, Cristina Sanches, Joana Pipa, and Leandro S. Almeida. 2016. "'To Be or Not to Be Retained … That's the Question!' Retention, Self-Esteem, Self-Concept, Achievement Goals, and Grades." *Frontiers in Psychology* 7 (October).

Rickles, Jordan H., and Michael Seltzer. 2014. "A Two-Stage Propensity Score Matching Strategy for Treatment Effect Estimation in a Multisite Observational Study." *Journal of Educational and Behavioral Statistics* 39 (6): 612–36.

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66 (5): 688–701.

Salza, Guido 2022 "Equally Performing Unfairly Evaluated- The Social Determinants of Grade Repetition in Italian High Schools", *Research in Social Stratification and Mobility*, forthcoming

Shepard, Lorrie A., and Mary Lee Smith. 1989. *Flunking Grades: Research and Policies on Retention*. Education Policy Perspectives. The Falmer Press, Taylor&Francis, Inc.
Stuart, Elizabeth A., and Donald B. Rubin. 2008. "Matching With Multiple Control Groups With Adjustment for Group Differences." *Journal of Educational and Behavioral Statistics* 33 (3): 279–306.
UNESCO Institute for Statistics 2012, *Global education digest 2012: opportunities lost: the impact of grade repetition and early school leaving*, UIS, Montreal.
West, Martin R. 2012. "Is Retaining Students in the Early Grades Self-Defeating." *CCF Brief* 49.
Wu, Wei, Stephen G. West, and Jan N. Hughes. 2008. "Short-Term Effects of Grade Retention on the Growth Rate of Woodcock–Johnson III Broad Math and Reading Scores." *Journal of School Psychology* 46 (1): 85–105.
Zhou, Xiang, and Yu Xie. 2020. "Heterogeneous Treatment Effects in the Presence of Self-Selection: A Propensity Score Perspective." *Sociological Methodology* 50 (1): 350–85.

APPENDIX

(A1) National descriptive statistics

The average share of retained students enrolled in Italian schools (all levels) was 6.6% in 2018, as compared to Piedmont (6.3%), Lombardy (6.8%), and Veneto (5.7%). As for the rest of the country, the share is 6.2% and 6.3% in the Center and the South, respectively, while it is higher in the islands, 8.5% (ISTAT 2018). The share of non-admitted students by grade has been declining over time, as Table A1.1 below shows.

Table A1.1. Share of retained students by grade in recent school years

	Middle school			High school				
Grades:	6th	7th	8th	9th	10th	11th	12th	13th
2013-2014	3.8%	3.2%	2.8%	16.3%	10.5%	9.7%	7.2%	4.2%
2014-2015	3.4%	2.1%	2.8%	14%	8.4%	7.8%	5.3%	4.6%
2015-2016	3.2%	2.7%	2.4%	14%	8.3%	7.8%	5.5%	4%
2016-2017	2.5%	2.1%	2.1%	13.4%	8.3%	7.6%	5.5%	3.8%

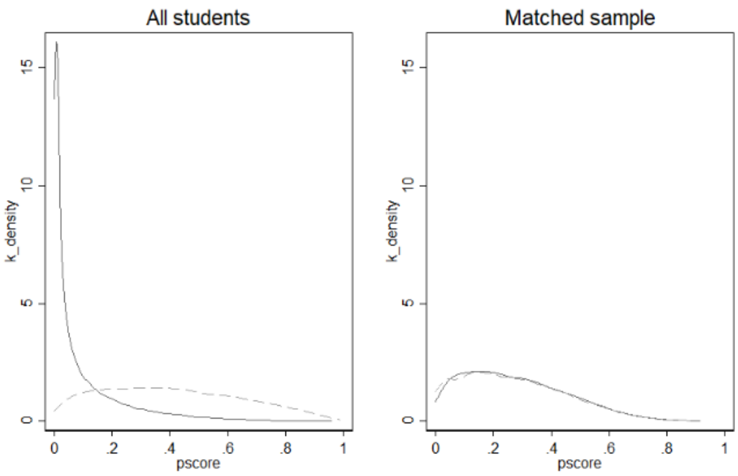
(A2) Analytic sample

Table A2.1. Steps to reach the analytic sample

	N
Enrolled students in 9th grade, 2014-15	179,350 -
Previously retained in 9th grade, 2013-14	23,285 -
Missing educational outcomes in 9th grade, 2014-15	2,864 =
Valid observations from ANS, 2014-15	153,201 -
Missing teachers' marks and\or INVALSI scores	29,856 =
Valid observation with performance info	**126,245**
N of retained in 9th grade, 2014-15	17,876
N of non-retained in 9th grade, 2014-15	108,369

(A3) Propensity score distributions

Figure 3: Propensity score distributions before and after matching



Solid line: promoted students
Dashed line: retained students.

(A4) Matching with different strictness to address the consequences of the violation of CIA

Let us formalize the problem.

Treatment assignment (at least in principle) depends on the following deterministic model:

$$T_{ij}^{*} = \beta_0 + \beta_1 A_{ij} + \beta_2 X_{ij} + \beta_3 Z_j + \mu_j \quad \text{(A4.1)}$$

$$T_{ij} = I(T_{ij}^{*} > 0)$$

where T^{*} is a retention score according to which the decision is taken and T is the binary treatment variable, Z captures school-level observed characteristics potentially affecting school learning standards (school curriculum and student body composition in terms of prior ability and family background). μ is a latent index of the school propensity to use retention as a remedial measure (*strictness*), assumed orthogonal to the other explanatory variables. Within each school, the pass or failure decision depends primarily on the students' current performance A, although other sociodemographic characteristics X may also play a role (Salza 2022).

Since we cannot include contemporary school performance and must rely on prior achievement measures, we may rewrite expression (A4.1) as:

$$T_{ij}^{*} = \beta_0 + \beta_1 (A_{-1ij} + \Delta A_{ij}) + \beta_2 X_{ij} + \beta_3 Z_j + \mu_j \quad \text{(A4.2)}$$

$$= \beta_0 + \beta_1 A_{-1ij} + \beta_2 X_{ij} + \beta_3 Z_j + \mu_j + \varepsilon_{ij}$$

where $\varepsilon_{ij} = \beta_1 \Delta A_{ij}$ is for the researcher an individual unobservable component.

Simplifying the setting, consider a binary outcome model, with Y_{ij}^* the propensity of one of the two options:

$$Y_{ij}^* = \gamma_0 + \gamma_1 A_{ij} + \gamma_2 X_{ij} + \gamma_3 Z_j + \theta T_{ij} + u'_{ij} \text{ (A4.3)}$$

In this specification, we assume that the outcome may depend on treatment and on the same observed characteristics affecting the grade repetition decision.

$$Y_{ij}^* = \gamma_0 + \gamma_1 (A_{-1ij} + \Delta A_{ij}) + \gamma_2 X_{ij} + \gamma_3 Z_j + \theta T_{ij} + u'_{ij}$$
$$= \gamma_0 + \gamma_1 A_{-1ij} + \gamma_2 X_{ij} + \gamma_3 Z_j + \theta T_{ij} + u_{ij} \text{ (A4.4)}$$

The error term $u_{ij} = \gamma_1 \Delta A_{ij} + u'_{ij}$ includes the contribution of new ability and the idiosyncratic error. Since the larger ΔA_{ij}, the lower the retention (treatment) probability, by neglecting the new ability component we will overestimate the negative impact of treatment or underestimate the positive effect of treatment.

<u>Proof</u>

We now demonstrate that by matching pupils in same-strictness schools we obtain the following:

$$E(Y^*|A_{-1}, X, Z, T = 1) - E(Y^*|A_{-1}, X, Z, T = 0) = \theta + \gamma_1 [E(\Delta A|\Delta A < k) - E(\Delta A|\Delta A > k)]$$
(A4.5)

In fact,

$$E(Y^*|A_{-1}, X, Z, T = 1) = \gamma_0 + \gamma_1 A_{-1} + \gamma_2 X + \gamma_3 Z + \theta + \gamma_1 E(\Delta A|T = 1) \text{ (A4.6)}$$

$$= \gamma_0 + \gamma_1 A_{-1} + \gamma_2 X + \gamma_3 Z + \theta + \gamma_1 E(\Delta A|\beta_0 + \beta_1 (A_{-1} + \Delta A) + \beta_2 X + \beta_3 Z + \mu > 0)$$

$$= \gamma_0 + \gamma_1 A_{-1} + \gamma_2 X + \gamma_3 Z + \theta + \gamma_1 E(\Delta A|\beta_1 \Delta A > -\beta_0 - \beta_1 A_{-1} - \beta_2 X - \beta_3 Z - \mu)$$

We can assume $\beta_1 < 0$, because the higher the performance the lower the retention probability; thus, the above expression becomes:

$$= \gamma_0 + \gamma_1 A_{-1} + \gamma_2 X + \gamma_3 Z + \theta + \gamma_1 E\left(\Delta A \middle| \Delta A < \frac{-\beta_0 - \beta_1 A_{-1} - \beta_2 X - \beta_3 Z - \mu}{\beta_1}\right)$$

Similarly:

$$E(Y^*|A_{-1}, X, Z, T = 0) = \gamma_0 + \gamma_1 A_{-1} + \gamma_2 X + \gamma_3 Z + \theta + \gamma_1 E(\Delta A|T = 0) \text{ (A4.7)}$$

$$= \gamma_0 + \gamma_1 A_{-1} + \gamma_2 X + \gamma_3 Z + \gamma_1 E\left(\Delta A \middle| \Delta A > \frac{-\beta_0 - \beta_1 A_{-1} - \beta_2 X - \beta_3 Z - \mu}{\beta_1}\right)$$

By subtracting (A4.7) from (A4.6) and letting $k = \frac{-\beta_0 - \beta_1 A_{-1} - \beta_2 X - \beta_3 Z - \mu}{\beta_1}$, we obtain (A4.5).

We may therefore conclude that if we match treated and control students with the same array of individual and school characteristics and the *same estimated strictness*, we will obtain a biased estimate of the true effect of grade retention. If Y^* is a positive outcome, $\gamma_1 > 0$, so the bias is $\gamma_1[E(\Delta A|\Delta A < k) - E(\Delta A|\Delta A > k)]$<0. If Y^* is a negative outcome, $\gamma_1 < 0$ and $\gamma_1[E(\Delta A|\Delta A < k) - E(\Delta A|\Delta A > k)]$>0. In either case, we would overestimate a negative effect or underestimate a positive effect.

What happens if we match each treated (retained) student with control (promoted) students in *less strict* schools? In this case, we end up estimating the effect of grade repetition θ as:

$$E(Y^*|A_{-1}, X, Z, T = 1) - E(Y|A_{-1}, X, Z, T = 0) = \theta + \gamma_1[E(\Delta A|\Delta A < k_T) - E(\Delta A|\Delta A > k_C)]$$

with $k_T > k_c$, because $k_T = \frac{-\beta_0 - \beta_1 A_{-1} - \beta_2 X - \beta_3 Z - \mu_T}{\beta_1}$ and $k_C = \frac{-\beta_0 - \beta_1 A_{-1} - \beta_2 X - \beta_3 Z - \mu_C}{\beta_1}$ and $\mu_T > \mu_C$ (keep in mind that $\beta_1 < 0$).

Since $E(\Delta A|\Delta A < k_T) - E(\Delta A|\Delta A > k_C) < E(\Delta A|\Delta A < k) - E(\Delta A|\Delta A > k)$, the bias, if still existing, would be smaller than with same-strictness matching. This means that if the effect of treatment is negative (i.e., grade retention worsens future outcomes), with less-strictness matching we reduce the bias attached to matching retained with non-retained students according to prior measures of ability. The larger the distance between the strictness levels of treatment and control groups, the smaller the bias, which tends to 0 as this distance increases.

Interestingly, since we actually end up matching retained students with higher prior school performance than the promoted ones, the bias could reverse direction, meaning that we could even deliver a conservative estimate of the impact of grade repetition.

(A5) Assumption check

Table A5.1. - Effect of estimated strictness on career type (unconditional outcomes).

Career	Exp($\boldsymbol{\beta}$) =RR	p-value
Linear (rif)	-	-
Downward	1.09	0.271
Grade repetition	1.13	0.028
Dropout	0.92	0.261

Multinomial logit models. Controls: gender, age, teacher marks, standardized scores, migration background, parental education. RR=relative risk of experiencing different career types (retained/promoted).

(A6) Sensitivity checks

Table A6.2. - Performance in 8th grade - different matching procedures

		Italian mark	Math mark	Behavior mark	Reading test	Math test
PS: range 0.20 bilateral. Strictness: the least (N=7453)						
Controls	mean	6.45	6.35	8.06	-0.45	-0.43
	sd	0.74	0.75	0.91	0.84	0.76
Treated	mean	6.62	6.45	8.32	-0.25	-0.27
	sd	0.76	0.76	0.8	0.71	0.71
Standardized Bias		0.22	0.13	0.33	0.28	0.23
PS: range 0.10 bilateral. Strictness: the least (N=5934)						
Controls	mean	6.48	6.35	8.12	-0.40	-0.40
	sd	0.72	0.71	0.86	0.74	0.82
Treated	mean	6.67	6.51	8.37	-0.23	-0.19
	sd	0.78	0.80	0.78	0.71	0.82
Standardized Bias		0.25	0.19	0.31	0.23	0.25
PS: range 0.20 unilateral. Strictness: the least (N=5528)						
Controls	mean	6.41	6.27	8.05	-0.46	-0.48
	sd	0.63	0.54	0.85	0.71	0.77
Treated	mean	6.68	6.52	8.38	-0.22	-0.18
	sd	0.78	0.81	0.79	0.70	0.82
Standardized Bias		0.34	0.32	0.42	0.35	0.36
PS: range 0.10 unilateral. Strictness: the least (N=4301)						
Controls	mean	6.49	6.34	8.14	-0.39	-0.37
	sd	0.68	0.62	0.83	0.71	0.78
Treated	mean	6.74	6.59	8.42	-0.17	-0.12
	sd	0.81	0.85	0.78	0.71	0.83
Standardized Bias		0.31	0.30	0.35	0.30	0.31
PS range 0.20 bilateral. Similar Strictness: ±5 percentiles (N= 4170)						
Controls	mean	6.66	6.56	8.34	-0.25	-0.19
	sd	0.85	0.91	0.82	0.75	0.89
Treated	mean	6.64	6.48	8.31	-0.27	-0.22
	sd	0.78	0.79	0.81	0.72	0.84
Standardized Bias		-0.03	-0.10	-0.04	-0.03	-0.04

Table A6.3. - LATT for different matching procedures

Career	Treated (%)	Controls (%)	RD	RR	p-value
PS: range 0.20 bilateral. Strictness: the least (N=7453)					
Linear	37.2	69.5	-32.4	0.5	0.000
Downward	20.6	2.2	18.4	9.2	0.000
Retention	16.0	23.9	-7.8	0.7	0.000
Dropout	26.2	4.3	21.8	6.0	0.000
PS: range 0.10 bilateral. Strictness: the least (N=5934)					
Linear	38.1	71.4	-33.2	0.5	0.000
Downward	20.8	1.8	19.0	11.5	0.000
Retention	16.0	22.7	-6.8	0.7	0.000
Dropout	25.0	4.1	21.0	6.1	0.000
PS: range 0.20 unilateral. Strictness: the least (N=5528)					
Linear	38.4	68.8	-30.4	0.6	0.000
Downward	21.0	2.3	18.8	9.3	0.000
Retention	15.8	24.5	-8.7	0.6	0.000
Dropout	24.8	4.5	20.3	5.5	0.000
PS: range 0.10 unilateral. Strictness: the least (N=4301)					
Linear	39.2	71.4	-32.2	0.5	0.000
Downward	21.4	2.4	19.0	8.8	0.000
Retention	15.9	22.2	-6.3	0.7	0.000
Dropout	23.6	4.1	19.5	5.8	0.000
PS range 0.20 bilateral. Similar Strictness: ±5 percentiles (N= 4170)					
Linear	36.8	73	-36.2	0.5	0.000
Downward	21.4	2.0	19.4	10.9	0.000
Retention	17.0	21.9	-4.9	0.8	0.000
Dropout	24.8	3.1	21.7	7.9	0.000

(A7) Heterogenous effects

Table A7.1. Interaction effects between grade repetition and parental education on the probability of a linear career. Logistic models, by school track.

	Coef.	S.E.	p-value	95% C.I.	
Traditional lyceums					
Treated	-2.11	0.16	0.000	-2.43	-1.79
Parents' highest education × grade repetition					
Diploma	0.04	0.19	0.826	-0.34	0.42
Tertiary	0.92	0.22	0.000	0.50	1.35
Other lyceums					
Treated	-1.65	0.21	0.000	-2.06	-1.24
Parents' highest education × grade repetition					
Diploma	-0.38	0.28	0.168	-0.92	0.16
Tertiary	-0.23	0.49	0.637	-1.19	0.73
Technical					
Treated	-1.59	0.07	0.000	-1.73	-1.44
Parents' highest education × grade repetition					
Diploma	0.26	0.11	0.013	0.06	0.47
Tertiary	0.19	0.27	0.469	-0.33	0.71
Vocational					
Treated	-1.36	0.10	0.000	-1.56	-1.15
Parents' highest education × grade repetition					
Diploma	-0.28	0.20	0.168	-0.67	0.12
Tertiary	-1.73	1.14	0.129	-3.96	0.51

Controls: gender, age, teacher marks, migration background, standardized scores. Ref. category is 'No diploma'.

Table A7.2. Interaction effects between grade repetition and migration background on the probability of a linear career. Logistic models by school track.

	Coef.	S.E.	p-value	95% C.I.	
Traditional lyceum					
Treated	-1.92	0.49	0.000	-2.88	-0.96
Migration background × grade repetition					
Citizen	0.07	0.49	0.878	-0.89	1.05
Other lyceum					
Treated	-1.04	0.99	0.294	-2.99	0.91
Migration background × grade repetition					
Citizen	-0.89	1.00	0.373	-2.85	1.07
Technical					
Treated	-1.56	0.23	0.000	-2.01	-1.11
Migration background × grade repetition					
Citizen	0.08	0.24	0.722	-0.38	0.55
Vocational					
Treated	-1.97	0.29	0.000	-2.54	-1.39
Migration background × grade repetition					
Citizen	0.54	0.30	0.077	-0.06	1.13

Controls: gender, age, teacher marks, migration background, standardized scores. Ref. category is 'Migrant'.