irvapp

The Effects of Remedial Exams on Student Achievement: Evidence from Upper Secondary Schools in Italy

Erich Battistin

Ilaria Covizzi

Antonio Schizzerotto

The Effects of Remedial Exams on Student Achievement: Evidence from Upper Secondary Schools in Italy

Erich Battistin
IRVAPP & University of Padova

Ilaria Covizzi
IRVAPP

Antonio Schizzerotto
IRVAPP & University of Trento

Istituto per la ricerca valutativa sulle politiche pubbliche
Fondazione Bruno Kessler
Via S. Croce 77
38122 Trento
Italy

Tel.: +39 0461 210242
Fax: +39 0461 210240
Email: info@irvapp.it
Website: http://irvapp.fbk.eu

Corresponding author: Antonio Schizzerotto, IRVAPP - Istituto per la Ricerca Valutativa sulle Politiche Pubbliche, Via S. Croce 77 - 38122, Trento, Italy. Email: schizzerotto@irvapp.it

The Effects of Remedial Exams on Student Achievement: Evidence from Upper Secondary Schools in Italy

March 2010

Abstract

We investigate the effect on student achievement of a remedial education reform recently introduced in upper secondary schools in Italy. Prior to its implementation, low-performing teenage students used to be given an 'educational debt' in one or more subjects. Such debt had to be recovered with no clear deadline or assessment during the student educational career. In response to increasing concerns about the low performance of Italian adolescent students, in 2007 the Italian Ministry of Education decided to re-introduce a remedial exam procedure already in place before school year 1993/94. This paper assesses whether a crucial feature of the reform, which is the threat of grade retention, has had any effect on student proficiency. To this end, we exploit the quasi experimental variation that results from geographical discontinuities in the implementation of the reform. Unlike the rest of the country, schools located in a well defined area in Northern Italy (province of Trento), which enjoys some degree of autonomy regarding education policies, opted out of the new progression system. We use this geographical variation to examine the effect of being at risk of grade retention on short-run achievement gains. We find positive effects on student achievement of children attending academic-oriented tracks but negative effects for students on technical/vocational tracks, thus suggesting a pattern of differential returns depending on the socio-economic background.

1.   Introduction

A wide empirical literature has addressed the determinants of student achievement over the last three decades. Large part of this literature has focused on the effects of school resources (Greenwald *et al.* 1996; Hanushek 1997; Wößmann 2003), class size and class composition (Angrist & Lavy 1999; Hoxby 2000; Dobbelsteen *et al.* 2002), peer effects (Sacerdote 2001; Markman et al. 2003), incentives to teachers (Lavy 2003), financial incentives to students and institutional differences between educational systems (Robin & Sprietsma 2003; Collier & Millimet 2009). There are only a few studies in which the effects of remedial education programmes for teenage students are investigated and selection bias issues are properly taken into consideration (Jacob & Lefgren 2004; Lavy & Schlosser 2005).

The purpose of this study is to assess the effectiveness of a remedial education reform recently introduced in the upper secondary school system in Italy. In particular, the paper examines whether a crucial feature of the reform, which is the threat of grade retention, has a positive effect on student proficiency.

Typically, an educational remedial programme is a course consisting in extra-class time offered to low-achieving students in order to improve their performance in one or more subjects. While remedial courses are not widespread in elementary and secondary education, they are quite common in post-secondary education (Alfred & Lum 1988; Boylan *et al.* 1999; Bettinger & Long 2007). For example, almost every college - especially community colleges - and university in the United States offers courses for students who are underprepared to take a particular course or pass a test. Yet, with a few exceptions (Calcagno & Long 2008), the literature provides little definitive evidence on the effectiveness of post-secondary remedial programmes.

With respect to remedial programmes for compulsory education systems, Jacob and Lefgren (2004) have investigated the effects on student achievement of a remedial intervention implemented in Chicago public schools, targeted to students in their third (aged 9), sixth (aged 12) and eight (aged 14) grades. The programme is part of an accountability policy aimed at ending the practice of passing students to the next grade regardless of their performance[1] by tying promotion to the achievement of minimum academic standards and, if necessary, to the attendance of remedial courses over the summer. Both remedial summer schools and grade retention are found to improve academic achievement of third-graders, but not of sixth-graders. Besides, since the identification strategy exploits a regression discontinuity idea, the external validity of the results is limited. Lavy and Schlosser (2005) have studied the impact of a remedial intervention for underperforming high school students in Israel designed to increase the percentage of those who earn matriculation certificates. Implementing a differences-in-differences strategy, the authors have found a significant increase in the school mean matriculation rate of participating students.

---

[1] Performance is measured through standardized high-stakes tests.

More generally, in recent years increasing concerns about the quality of education have lead to the implementation of accountability policies designed to hold administrators, teachers and students responsible for the level of student achievement both in Europe and the Unites States (OECD 2007a). Educational accountability strategies reflect the belief that "the promise of rewards or the threat of sanctions is needed to ensure change"(Hamilton 2003: 35). Thus, accountability systems not only set standards and tools to assess student performance, but also provide rewards and sanctions for schools or districts based on such performance assessment.

Although strictly speaking the intervention we consider in what follows is not an accountability measure, it shares with accountability policies the central assumption that punishments are an effective tool to enhance student performance. To this end, we exploit the quasi experimental variation that results from geographical discontinuities in the implementation of the reform. Unlike the rest of the country, schools located in a well defined area in Northern Italy (the administrative province of Trento), which enjoys some degree of autonomy regarding education policies, opted out of the new progression system. We use this geographical variation to examine the effect of being at risk of grade retention on short-run achievement gains. The co-presence of a "treatment" group (students subject to new progression rules) and a quasi experimental "control" group (students undergoing the old rules) allows us to assess the causal effect of the reform, and to shed light on whether, and to what extent, it has worked as a tool to enhance student achievement. Student outcomes are measured through an assessment test constructed *ad-hoc* for this study and largely based on the *OECD Programme for International Student Assessment* (PISA) test. Unlike previous studies, the identification strategy employed here yields causal conclusions whose validity holds across an heterogeneous group of students, not only for low-achieving students.

The remainder of the paper is structured as follows. Features and background of the policy under evaluation are reported in Section 2; Section 3 describes the evaluation design; data and methods employed are illustrated in Section 4; results are reported in Section 5; conclusions and policy implication are discussed in Section 6.

2. Background

Starting from school year 2007/08, upper secondary schools in Italy[2] are required by the Ministry of Education to implement remedial education programmes to help low-achieving students. Until school year 2006/07 students who did not meet proficiency

---

[2] Both primary and secondary education in Italy is mainly provided by public schools. Upper secondary education (*scuola superiore di secondo grado*) normally last 5 years (from 14 to 19 years of age). In the upper secondary system, types of schools are differentiated by subjects and activities: the *licei* (with a specifically academic curriculum), the *Istituti tecnici* (Technical institutes) and the *Istituti professionali* (Vocational institutes). For more detailed information on the Italian education system, see for example Eurydice 2009.

expectations were given a *debito formativo* ('educational debt'), that is a final mark signaling student's failure in one or more subjects. Such lack in achievement was to be recovered in the following years with no clearly defined deadline, resulting in a *de facto* 'social promotion' practice passing students to the next grade regardless of their performance. With the new progression system, low-performing students are compelled to recover their educational lacks by passing a remedial exam before the beginning of the new school year[3]. On failing to do so, students face the risk of grade retention. Students who obtain at least 6/10 are admitted to the following grade. If marks lower than 6/10 persist, retention will be deliberated by the classroom council. However, due to school autonomy, the standards to be achieved and assessment instruments are not set by the State or school districts as in standard-based accountability policies, but are proposed by the teachers and approved by the classroom council, which is ultimately responsible for student promotion and retention . As a consequence of school autonomy, such exams can be either written or oral or practical sessions (labs) depending on the classroom council.

Remedial exams were first introduced in the Italian school system in 1923 by the so called "Gentile reform". In the late 1980s they were abolished in primary and lower secondary schools, and in the early 1990s they were suppressed also in upper secondary schools. The policy rationale for their reintroduction in 2007 resulted from a combination of scientific and political debates on the low performance of Italian students compared to their peers in other advanced countries. Although the poor learning effectiveness of the Italian upper secondary system and the need for a general reform of the curricula had been previously brought to attention,[4] it was the disappointing output on reading literacy - documented by the 2000 PISA survey - that first generated a major concern in the public opinion. In fact, Italy ranked above Spain, Portugal and Greece but far behind most advanced countries. The average score of Italian students was 100 points lower than that of top-ranking Korean students (OECD 2001). The public concern became widespread after the 2003 PISA survey results on mathematics, when the overall performance of Italian students dropped below that of Spain and Portugal with an average score of 86 points lower than that of their Finnish counterparts (OECD 2004). The overall picture was confirmed in 2006 with the PISA outcomes on scientific literacy (OECD 2007a). Moving from this evidence, it was decided in 2007 to reintroduce remedial exams for low-achieving students (the so called "Fioroni reform") and to make remedial courses for such students compulsory. As a result, starting from the school year 2007/08 low-performing students failing to pass the exam face the risk of grade retention.

Contrary to what happened in the rest of the country, the local government of the province of Trento has refused to comply with such reform. PISA scores for students

---

[3] "At the end of each school year, class teachers at the class council assign final marks to each student (*scrutinio*). Marks are proposed by teachers of each subject to the classroom council and approved by the majority of teachers. If no majority is reached, the vote of the Headmaster prevails. Admission to the next grade requires marks equal to or higher than 6/10 in each subject." (Eurydice 2009: 117).
[4] See for example Bottani (1986), Benadusi and Consoli (2004), Cobalti and Schizzerotto (1993), Gasperoni (1996). The Italian school system exhibited a low-performance profile both in the 1995 and 1999 sessions of *Trends in International Mathematics and Science Study* (Timms), the survey carried out every four years on the mathematics and science achievement of students from primary and lower secondary schools.

living in that province   are at least as good as those recorded by top-ranking countries.[5] In light of this evidence, local policy-makers have decided that there is no real need to comply with the national intervention. Furthermore, they support the idea that remedial courses already offered to students in that area are sufficient to help them repay their 'educational debt', and that no extra exam before the beginning of the new school year is needed to guarantee the achievement of the required academic standards.

Reactions to the policy were mixed and generated a heated debate on the merits and demerits of the intervention. On the one hand, advocates of the intervention view remedial exams as a positive punitive tool to (i) convince students to raise their commitment in learning *all* school subjects (also those usually considered as marginal components of their curricula) as a consequence of the threat of grade retention, and (ii) persuade teachers not to be too tolerant towards their students. On the other hand, opponents have raised the concern that remedial exams may set up low-achieving students to fail and the threat of grade retention may undermine effort and engagement, especially of struggling students. Put differently, remedial exams may contribute to increase school selectivity and, therefore, enhance inequalities in educational opportunities.

As a matter of fact, advocates of the reform believe that the threat of repeating a grade is a device to limit undesired behaviour - that is, low school performance. According to this interpretation, students are supposed to act in ways which reduce the threat of punishment, the latter being grade retention. In other words, the threatened student is expected to study more intensely and, as a consequence, to achieve higher levels of proficiency because individuals instinctively fair failure. From a theoretical point of view, such approach echoes the behaviourist theories of learning. In fact, the assumption underlying such view broadly falls within the reinforcement theory originally developed by the behaviourist school of psychology (Skinner 1938; Staddon 2003). However, the 'stimulus-response' mechanism alone is not sufficient to account for all outcomes observed in learning situations. For example, according to cognitivism, students are not 'black-boxes' reacting to external stimuli but individuals who require active participation in order to learn, and whose actions are a consequence of thinking, not simply a response to an external stimulus. As we will see in the results section and as shown in previous empirical works, intrinsic motivation (Kruglanski 1978; Deci *et al.* 1999; Roderick & Engel 2001; Benabou & Tirole 2003), social origins (Ishida *et al.* 1995; OECD 2001; Erikson *et al.* 2005; Jackson *et al.* 2007), parents' and teachers' expectations (Bronstein *et al.* 2005; Spera *et al.* 2009; Mistry *et al.* 2009), and teacher's classroom assessment practices (Hamilton 2003; Allensworth 2005) are also likely to play a pivotal role in the learning process and produce differentiated students' reactions to punishment practices.

---

[5] It is worth noting that Italy is characterized by wide variation in PISA scores across areas and school tracks. Upper secondary school students in the Northern regions perform far better than the national OECD average; while those residing in the Southern regions of the country perform as poorly as students living in developing countries. Moreover, controlling for area of residence, Italian students attending academically oriented schools (*Licei*) perform better than those attending either technically oriented (*Istituti tecnici*) or vocationally oriented (*Istituti professionali*) schools (OECD 2009).

3. Data

We exploit unique data on student-level standardized test scores. The data set combines school administrative databases (containing teachers' marks and information on promotion/retention) and data from two different sample surveys purposively designed for this study. The first survey collects information on student proficiency through the administration of a standardized assessment test to all students in our sample[6]. The second survey – administered to parents of sampled students - collects information on parental social background such as education, job status, household composition, learning resources at home.

The assessment test is based on the well-known OECD-PISA measure, yet adjusted to the specific purpose of our study. PISA measures how well students near the end of compulsory education (aged 15) are able to apply their acquired knowledge to problems related to real-life situations. Unlike other assessment tests (e.g. TIMSS, PIRLS, etc.), PISA does not provide a direct measurement of attained curriculum knowledge. In PISA the three domains of reading, mathematical, and scientific literacy are assessed through a combination of multiple-choice items and questions requiring students to construct their own responses[7]. The test administered to our sample is a short version of the original PISA test and is based on publicly released items from the first three assessments (PISA 2000, 2003, 2006). The way the test was constructed guarantees comparability with the official PISA scale, in that the structure of item difficulty was maintained coherent with the original test. The items were presented to students in three one-hour booklets resulting in a three-hour session for each student; whereas in the OECD-PISA survey a total of about seven hours of test items is covered. There were 23 reading units, 20 mathematics units and 19 science units. Unlike the official PISA survey, all students in our sample took the same tests, not a different combinations of test items. Following the OECD procedure, test scores were obtained employing scaling models based on *item response theory* (IRT) methodologies (Masters, 1982). To allow for comparability with PISA 2006, scores have been re-parameterized on mean and standard deviation of PISA 2006 scores in the province of Trento. At the end of the test administration, students were also asked to answer a 'context questionnaire', which took 15 minutes to complete, providing information about their parents' education and jobs, and their life-style at home.

An additional survey was carried out on parents. They were asked to provide detailed information on their educational and employment background, household composition and home learning resources. Lastly, information on student achievement was complemented with administrative data on teachers' marks on past years at school as well as with the final grade students obtained at the state examination on completion of the lower secondary school (leaving certificate).

---

[6] The choice of a standardized test relies on the evidence that teachers' marks are not an appropriate measurement of student achievement as they may reflect subjectivity (Allensworth 2005; Hamilton 2003).

[7] At each administration (every 3 years) one subject is assessed in depth on a rotating basis.
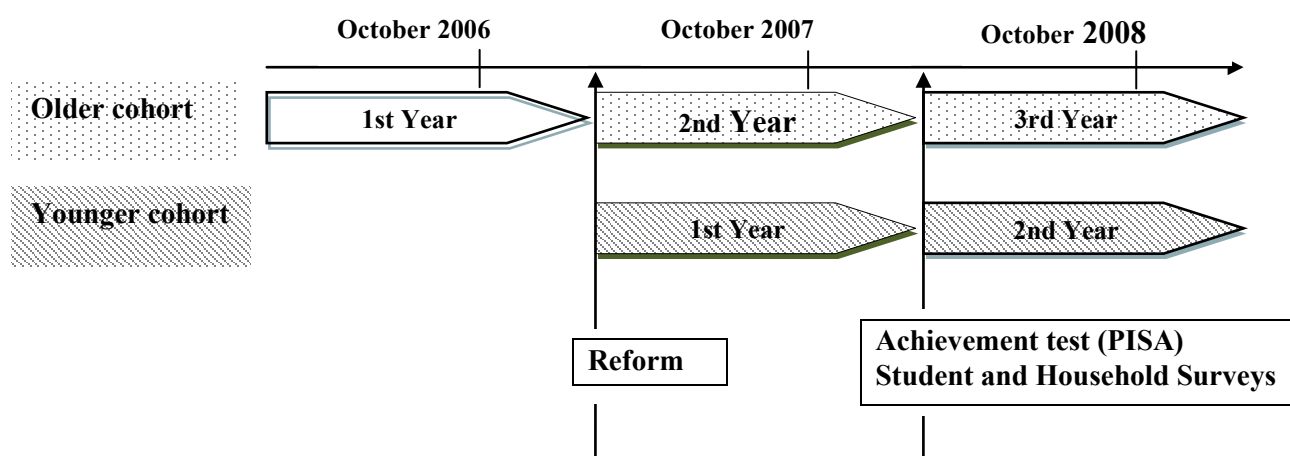
3.1   Sampling frame

To construct the sampling frame we started by considering a selected number of towns sharing similar characteristics in terms of their demographic, economic and occupational structure as well as of school-related infrastructures. Being the province of Trento the only area where the 'educational debt' is still maintained, we chose towns along the administrative borders of that area. The leading criteria followed to guide selection of towns were i) the presence of schools for each track of the Italian upper secondary school system: *licei* (academic, or general education, track), i*stituti tecnici* (technical track) and *istituti professionali* (vocational track); ii) population size of town; and iii) features of the economic and occupational structure. We then followed a *one-to-one* matching procedure, and selected for three towns of the control group (Trento province) the most similar counterparts along the administrative border. Pair-wise matching was conducted by using the same criteria, that were further refined controlling for geographical proximity (less than seventy kilometers).

The target sample resulted from a two-stage procedure that selected schools in the first stage, and in the second stage cohorts of students defined from the year attended in October 2008. We again followed a *one-to-one* matching procedure, selecting similar schools located in each pair of towns. The selection of schools was conducted by controlling for observable dimensions such as school track, school size as measured by trends in enrollment and school resources, as well as unobservable dimensions (such as reputation of the school) gathered from general knowledge of the socioeconomic background in which they operate.

Across all schools, we focused on students attending the second and the third year during the school year 2008/09, thus aged between 15 and 16. For each school we randomly selected two classes in the second year (i.e. for the cohort of students enrolled for the first time in school year 2007/08) and two classes in the third year (i.e. for the cohort of students enrolled for the first time in school year 2006/07). As we shall see, such a design ensures variability in the duration of enrollment at school across the different regimes defined by the reform (see fig.1).

To summarize, our sample comprises students aged 15-16 in similar schools located in three selected pairs of towns located either sides of the administrative border of the Trento province. The number of schools involved in the analysis is 22, of which 11 in the Trento province and 11 outside. The number of students in the younger and the older cohorts is 863 and 827, respectively, and these numbers are roughly equally distributed in the two school tracks (academic and vocational/technical). Table 1 presents the sample size relative to students, separately by treatment/control area, student cohort (2006/07 and 2007/08) and school track (general/academic and technical/vocational).

Fig. 1 *Sample design of student cohorts*

To all students in the sample we administered the standardized assessment test in the period October/November 2008 (right after the beginning of the school year 2008/09) along with the questionnaire collecting demographics and family background information. For students in the older cohort (2006/07) we were able to retrieve information on teachers' marks during the first and the second year at school, as well as on the final outcome (passed or not passed to the next grade) at the end of each school year. For students of the younger cohort (2007/08), we were able to obtain the same information relative to their first year at school.

Tab. 1 *Sample size by treatment/control area, student cohort and school track.*

	Treatment areas (remedial exam)			
	Student cohort 2006/07		*Student cohort 2007/08*	
	Academic schools	*Technical/vocational schools*	*Academic schools*	*Technical/vocational schools*
Students	230	229	209	215
Classes	11	14	10	14
Schools	4	7	4	7

	Control areas (educational debt)			
	Student cohort 2006/07		*Student cohort 2007/08*	
	Academic schools	*Technical/vocational schools*	*Academic schools*	*Technical/vocational schools*
Students	152	216	189	250
Classes	10	14	10	14
Schools	4	7	4	7

Note: *"Control areas" were selected within the province of Trento, while "treatment" areas were selected from neighborhood provinces.*

Table 2 summarizes the information on student performance available for the two cohorts of students across school years *before* and *after* the remedial exam reform. The first year at school for the older cohort of students (those enrolled in 2006/07) refers to a *pre-reform* period, and thus performance during this year is not affected by the reform (under the reasonable assumption of no anticipatory effects). The second and the third year at school are instead referred to *post-reform* periods, so that the observed outcome would in general depend on whether the school attended is inside or outside the Trento province. The standardised measure taken at the beginning of the third year will therefore depict student proficiency after one year from the rolling out of the reform. As for the younger cohort of students (those enrolled in 2007/08), no information is collected at upper secondary school in a pre-reform period. In other words, no variability in policy regime can be exploited for this cohort, and the only measurement of performance available is the final grade at lower secondary level. In our empirically analysis we will pursue a number of strategies to retrieve the causal effect of the reform for this group using the information available.

Tab. 2 *Information available on student performance (teachers' marks and assessment test) at secondary school by student cohort. Information available for post-reform periods are in grey.*

	Older student cohort (enrolled in 2006/07)	Younger student cohort (enrolled in 2007/08)
Final grade at lower secondary leaving certificate *(Source: school administrative database)*	X	X
Teachers' marks at 1st semester (1st year) *(Source: school administrative database)*	X	X
Teachers' marks at 2nd semester (1st year) *(Source: school administrative database)*	X	X
Teachers' marks at 1st semester (2nd year) *(Source: school administrative database)*	X	
Teachers' marks at 2nd semester (2nd year) *(Source: school administrative database)*	X	
Assessment test *(Source: survey carried out in October 2008)*	X	X

4. Evaluation design

The aim of this section is twofold. First, we will discuss the conditions required to identify the causal effects of the introduction of remedial exams on student achievement. Second, we will discuss the estimation approach implemented to derive the results presented in Section 5.

4.1 Identification strategy

In its bare essentials our identification strategy exploits the *quasi experimental* variation induced by geographical discontinuities in the implementation of the policy. As discussed above, remedial exams were introduced countrywide starting from the school year 2007/08, with the exception of the Trento province. The evaluation design sets up the comparison of outcomes for students in upper secondary schools in the Trento province (control group) to outcomes for similar schools in adjacent areas (treatment group). Outcomes are referred to student achievement measured through the assessment test presented in Section 3.

The *causal* interpretation of differences in achievement observed between schools inside and outside the Trento province crucially relies on a *ceteris paribus* condition about the composition of students in the two groups of schools. This amounts to assuming that the outcome for students enrolled in one group of schools can serve as an approximation to the *counterfactual* outcome for students enrolled in the other group of schools. That is, school outcomes of students in the control group should closely resemble what students in the rest of Italy would have experienced, had remedial exams not been introduced.

The general problem underlying the validity of this condition can be easily put across using standard arguments taken from the programme evaluation literature (see, for example, Rubin 1974, and Heckman *et al.* 1999). In the *potential outcomes* framework interest lies in the causal impact of a given 'treatment' on an 'outcome' of interest. In the following, let the 'treatment' be the remedial exam and let the 'outcome' be school achievement. Let Y_1 (Y_0) denote the *potential* outcome that would result from the remedial exam being (not being) in operation. The causal effect of the reform on school achievement is then defined as $Y_1 - Y_0$, and corresponds to the difference in achievement induced by the introduction of remedial exams. Note that this difference is by its very nature not observable, as geographical location of the school attended reveals only one of the two potential outcomes (Y_0 for students in the Trento province, and Y_1 otherwise).

The *average* policy impact for students facing remedial exams (or the *average treatment effect of the programme on the treated*, ATT) is defined as:

$$E[Y_1 - Y_0|D = 1] = E[Y_1|D = 1] - E[Y_0|D = 1],$$

where D denotes a dummy variable for schools outside the Trento province. Throughout our discussion the ATT will represent the causal parameter of interest. The evaluation problem consists of dealing with the missing data problem that precludes direct estimation of $E[Y_0|D = 1]$: data are only informative about (features of) the distribution of Y_0 for $D = 0$ schools, and about (features of) the distribution of Y_1 for $D = 1$ schools. This term refers to a *counterfactual* situation which is not observable in the data, requiring as it does knowledge of what the average achievement would have been in schools outside the Trento province, had remedial exams not been introduced.

The estimators used in this paper rely on assumptions that allow to retrieve the missing counterfactual term. The key econometric difficulty results from the non-random selection of students into schools. For example, better students are more likely to sort into better schools, and the quality of students with respect to their cognitive abilities

developed at primary school may vary geographically a great deal. Under the assumption that conditioning on an available set of covariates X pre-determined with respect to the implementation of the reform removes all systematic differential selection, one could retrieve the counterfactual term of interest. Under this condition, the counterfactual term can be written as:

$$E[Y_0|D = 1] = \int E[Y_0|D = 0, X = x]f(x|D = 1)dx$$

so that the causal parameter of interest can be backed out from the raw distributions of achievement and of the X's in the two groups of schools (Heckman *et al.* 1999).

The last expression embodies the idea of a *conditional independence assumption* which Heckman and Robb (1985) call "*selection on observables*". This amounts to assuming that compositional differences in the pool of students attending $D = 0$ and $D = 1$ schools are solely limited to *observable* characteristics. Thus, the validity of such assumption in the context of this study crucially boils down to knowledge of the determinants of cognitive achievement that are relevant to define heterogeneity in the composition of schools in the two groups of areas.

An extensive and multidisciplinary literature has studied over the years the relationship between schooling inputs and test score outcomes for school-age children. Researchers have often drawn an analogy between the knowledge acquisition process and the production process of a firm to frame the problem within the context of a production function for cognitive achievement. The general framework for modeling the education production process laid out in Todd and Wolpin (2003) is consistent with child development being a *cumulative* process depending on the full history of inputs applied by families and schools as well as on children's inherited endowments (such as innate ability). Thus, the extent to which compositional differences are important for the assessment of the causal relationship addressed in this paper needs to be carefully discussed in light of the information available in the data.

It follows from the discussion above that, in analyzing cognitive achievement, the set X should include lifetime history of family and school inputs as well as on heritable endowments of students. Assessing the causal effect of introducing remedial exams via a conditional independence assumption is challenging if some of these components are *not* observable, for example because of incomplete input histories and/or unobserved heritable endowments.

Two important points in this respect are worth making. First, to reduce the degree of compositional differences between school inside and outside the Trento province, a "matched pilot/control" design for schools involved in the analysis was implemented (see Section 4 for further details). We then limited our analysis to students in these two groups of schools, thus controlling for the extent of heterogeneity across students on the one hand, but admittedly paying in terms of *external validity* of our results on the other. The *internal validity* of the design is strengthened by its similarities with a *regression discontinuity* strategy, the discontinuity holding with respect to the administrative border of the Trento province.

Second, as it will be clear from the next section, data collected for this study contain information on schooling inputs at one point in time, and contain only contemporaneous family background information. The lack of historical information on schooling and household inputs (and on innate ability) is a situation that is often faced by empirical studies that use the education production literature (see, for example, Todd and Wolpin 2003, for a review). This limitation has led us to balance the distribution of the two groups of students with respect to characteristics that are included in a *value-added* specification of the education production function. Achievement at upper secondary school is modeled through a large set of contemporaneous school and family input measures and lagged (baseline) achievement at the beginning of the first year. Standardized tests will be used to measure cognitive progression at school, whereas administrative information on teachers' marks will provide the basis to control for compositional differences with respect to ability at enrollment.

4.2　Estimation

Estimation will assume throughout that, netting off the effect of X, the mean outcome comparison for students enrolled in schools either side of the border identifies the causal effect of remedial exams on achievement. We will proceed by allowing such effect to be *heterogeneous* across students, and will draw inference based on models that are robust to the particular specification adopted. We will consider estimator obtained as empirical analogues of the following two expressions for the ATT:

$$\int \{E[Y|D = 1, X = x] - E[Y|D = 0, X = x]\} f[x|D = 1] dx, \tag{1}$$

$$\int \{E[Y|D = 1, e(x)] - E[Y|D = 0, e(x)]\} f[e(x)|D = 1] dx, \tag{2}$$

which both holds under the selection on X, being $e(x)$ the *propensity score* (see Rosenbaum and Rubin, 1983).

In particular, we will report values of the average return obtained through the following procedure. First, we will consider the regression:

$$y_i = \beta_0 + \beta_1 d_i + \beta_2 d_i x_i + \beta_3 x_i + \varepsilon_i,$$

which models heterogeneity in individual returns through the full set of observables X. According to this specification there is:

$$E[Y|D = 1, X = x] - E[Y|D = 0, X = x] = \beta_1 + \beta_2 x,$$

which if used in (1) implies:

$$E[Y_1 - Y_0|D = 1] = \beta_1 + \beta_2 E[X|D = 1].$$

The estimated effect is then obtained by considering the empirical counterpart of the last expression:

$$\hat{\beta}_1 + \hat{\beta}_2 \frac{1}{n_1}\sum_{i=1}^{n_1} x_i, \tag{3}$$

being n_1 the number of students in $D = 1$ schools and $\hat{\beta}_1$ and $\hat{\beta}_2$ the sample counterparts of the regression coefficients.

The sensitivity of results with respect to the parametric specification adopted is investigated by considering the semi-parametric counterpart of (2) implemented via matching on the propensity score $e(x)$. In particular, we studied the importance of *common support* problems in the data and their effects on estimation (see Black and Smith, 2006). The estimated effect is obtained from the quantity:

$$\frac{1}{n_1}\sum_{i=1}^{n_1}(y_i - \hat{y}_i),$$

where:

$$\hat{y}_i \equiv \frac{1}{n_0}\sum_{j=1}^{n_0} K\{|e(x_i) - e(x_j)|\}y_j, \tag{4}$$

being n_0 the number of students in $D = 0$ schools and $K\{|e(x_i) - e(x_j)|\}$ a weight given to the j-th student in $D = 0$ schools which depends on the distance on her propensity score $e(x_j)$ and the propensity score $e(x_i)$ of the i-th student in $D = 0$ schools. Alternative weights define a alternative matching estimators of the average return.

5. Results

5.1 Descriptive statistics

Table 3 provides a picture of the degree of homogeneity for students in treated schools and control schools. It presents means and standard deviations of covariates used later on to compute the treatment average effect (see Section 5.2). Covariates include key students' socio-demographic characteristics as well as education, labour market participation, socio-economic status of parents[8] and life-style deprivation of the household[9]. Also reported are descriptive statistics of teachers' marks that refer to the time spell starting from the first year of the upper secondary school cycle up to the

---

[8] The socio-economic status is measured using an Italian occupational stratification scale that measures the social desiderability (and, in broad sense, the prestige) attributed to different jobs (de Lillo & Schizzerotto 1985).

[9] The life style deprivation index (Whelan *et al.* 2002) is an additive index based on the lack of 5 items in the household: TV, car, DVD player, computer, internet access. Each individual item is weighted by the proportion of households possessing that item in Italy. Weights derive from IT-SILC 2006.

administration of the assessment test[10]. As a result of our sampling design, the distributions of covariates for the treatment and control groups are expected to be roughly the same. Columns 3 and 4 of table 3 show that both the treatment and control groups exhibit very similar distributions.
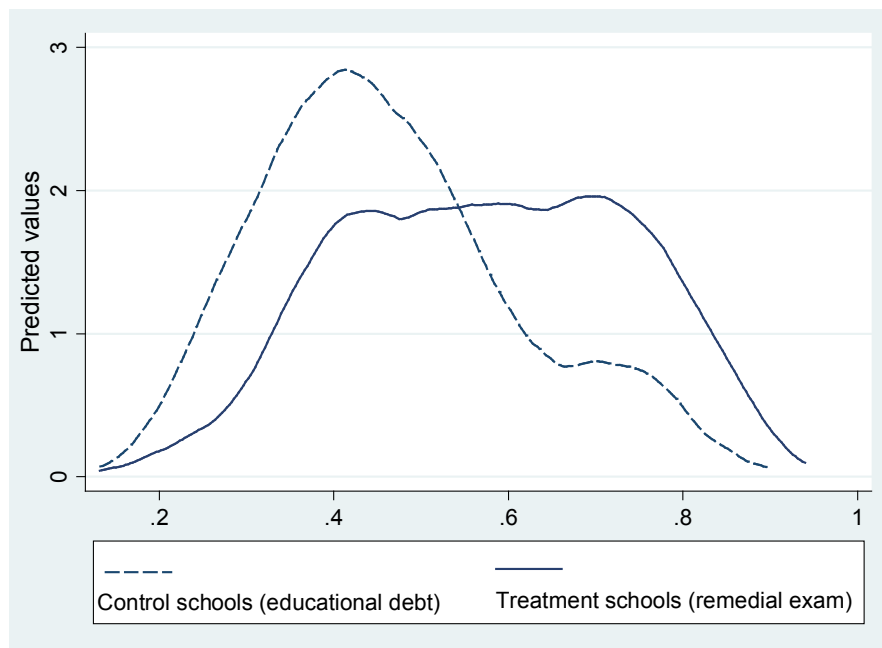
Tab. 3 *Descriptive statistics for the student sample*

Variables	Source	Treated schools		Control schools		Probit Regression	
		Mean	Sd	Mean	Sd	Coef	p-value
Student Demographics							
Gender (male)	SQ	0.47	0.50	0.48	0.50	-0.01	0.700
Age	SQ	15.92	0.75	15.87	0.75	-0.00	0.595
Dummy for foreign student	SQ+HQ	0.33	0.47	0.25	0.44	0.09	0.004
Dummy for cohabitation with mother and father	SQ	0.84	0.36	0.86	0.35	-0.07	0.096
School in the place of residence	SQ+AD	0.39	0.49	0.36	0.48	0.06	0.034
Household Demographics							
Father's age	SQ+HQ	48.56	5.64	48.39	5.30	-0.00	0.571
Father's education	SQ+HQ						
Primary		0.28	0.45	0.27	0.45		
Secondary		0.56	0.50	0.53	0.50	-0.01	0.829
Tertiary		0.17	0.36	0.20	0.40	-0.02	0.659
Mother's age	SQ+HQ	45.07	4.80	45.23	4.87	-0.00	0.713
Mother's education	SQ+HQ						
Primary		0.28	0.45	0.22	0.41		
Secondary		0.54	0.50	0.59	0.49	-0.09	0.004
Tertiary		0.18	0.38	0.19	0.39	-0.15	0.003
Dummy for housewife mothers	SQ+HQ	0.19	0.39	0.20	0.40	-0.01	0.747
Dummy for unemployed mother or father	SQ+HQ	0.02	0.16	0.02	0.14	-0.04	0.610
Household Wealth and Social-Status							
Occupational prestige scale	HQ+SQ	45.77	19.34	46.96	20.07	-0.00	0.362
Life-style deprivation index	SQ	0.40	1.26	0.44	1.30	0.00	0.782
Student's Career							
School type (vocational or academic)	AD						
- Academic schools						0.04	0.235
- Technical/Vocational schools							
Year attended at interview (2nd or 3rd)	AD						
- 2nd		0.48	0.50	0.54	0.50	-0.07	0.015
- 3rd		0.52	0.50	0.46	0.50		
Final grade at lower secondary leaving certificate (8th grade)	AD						
- Sufficient		0.17	0.37	0.22	0.42		
- Good		0.27	0.44	0.30	0.46	0.04	0.407
- Very good		0.27	0.44	0.30	0.45	0.04	0.392
- Excellent		0.30	0.46	0.19	0.39	0.19	0.001
Grade at the end of the first semester (1st year at secondary school): Italian literature	AD	6.45	0.90	6.48	0.81	-0.08	0.000
Grade at the end of the first semester (1st year at secondary school): Math	AD	6.59	1.21	6.37	1.08	0.07	0.000
Grade at the end of the first semester (1st year at secondary school): *Root-specific subjects*	AD	6.58	1.01	6.59	0.97	-0.04	0.043
Grade at the end of the first semester (2nd year at secondary school): Italian literature - only students at the 3rd year in 2007/08	AD	6.45	0.87	6.38	0.85		
Grade at the end of the first semester (2nd year at secondary school): Math - only students at the 3rd year in 2007/08	AD	6.45	1.24	6.22	1.09		
Grade at the end of the first semester (2nd year at secondary school): Root-specific subjects - only students at the 3rd year in 2007/08	AD	6.55	1.02	6.40	0.94		
Number of observations		*883*		*807*		*1690*	

---

[10] That means that figures for the second year include only marks for the older cohort of students, and that the set of variables common across the two cohorts refers only to the final grade at lower secondary school and to teacher's marks during the first year at upper secondary school.

To control for the effectiveness of our sampling procedure, we specified a binary regression model (*probit*) for the probability of being a student in the treatment schools (schools outside the province of Trento) as opposed to being a student in the control group. The covariates considered are those listed in table 3. Results are reported in the last column of table 3. Coefficients show that, despite few dimensions considered[11], none of the above covariates substantially affect the outcome variable.

To control for the lack of common support issues, we have plotted in figure 2 the distribution of predicted values from the above illustrated binary regression along with the distribution of predicted values from the binary regression in which the outcome variable is a dummy for being a student in the treated schools. The figure provides graphical evidence that rules out the common support problems in the data as, with few exceptions, all values of X are observed amongst treated and non-treated. On the whole, for each student in the treatment group there is a student with similar characteristics in the control group that can be used as a matched comparison observation.

Fig. 2   *Distributions of predicted values of the probability of being a student in the treatment vis-à-vis control schools*

---

[11] Students in the treatment group are *ceteris paribus* slightly more likely to be foreigners and to have mothers with lower educational attainment compared to the control group. Differences in the distribution of past student achievement are also noticeable. In particular, the performance at the final examination for the lower secondary leaving certificate is skewed towards lower values for students in the control group, who instead performed better in their first year of upper secondary school when compared to their peers in the treatment group.

In table 4 we turn our attention to the outcomes of our assessment tests for reading, maths and science. The breakdown is presented for average scores by cohort of enrollment in the first year, school track and area. The average score for the three domains considered is considerably lower for students enrolled in vocational or technical schools compared to students in academic schools. Not surprisingly the distribution of scores is skewed towards larger values for the older cohort of students. As it can be seen from the width of confidence intervals, there are statistically significant differences between the two cohorts along all the three dimensions considered but this differences disappear across areas within the same cohort.

Tab. 4 *Standardized test mean scores in the discipline areas of reading, mathematics, and science by student cohort and type of school track.*

Academic schools

	Student cohort 2006/07						Student cohort 2007/08					
	Outside Trentino			Trentino			Outside Trentino			Trentino		
	Mean	[95% conf int.]		Mean	[95%conf int.]		Mean	[95% conf int.]		Mean	[95% conf int.]	
Reading	526.67	513.96	539.39	528.94	513.08	544.81	501.43	488.73	514.13	497.04	482.43	511.65
Mathematics	520.15	509.15	531.15	515.66	501.64	529.67	506.16	493.18	519.14	496.44	483.15	509.73
Science	535.35	524.38	546.33	527.89	514.74	541.04	524.71	512.47	536.96	508.15	495.74	520.57
N	230			152			209			189		

Technical/Vocational schools

	Student cohort 2006/07						Student cohort 2007/08					
	Outside Trentino			Trentino			Outside Trentino			Trentino		
	Mean	[95%conf int.]		Mean	[95%conf int.]		Mean	[95%conf int.]		Mean	[95%conf int.]	
Reading	426.94	417.34	436.92	448.97	435.89	461.71	396.19	384.61	407.34	418.56	408.58	428.83
Mathematics	476.66	464.08	489.15	473.55	458.26	488.10	433.05	416.03	450.17	439.94	426.53	454.01
Science	456.51	444.69	468.69	458.10	445.01	470.75	422.44	408.65	435.87	455.36	443.85	467.23
N	229			216			215			250		

In table 5 average teachers' marks in maths by tertiles of test scores are reported. In particular, we consider teachers' marks referred to the end of the 2007/08 school year, which is to the closest point in time to when the test was administered. As expected, average marks monotonically increase with increasing tertiles of test scores. Conditional on tertiles of test scores, average marks in vocational schools are systematically below those in academic schools. However, as documented in several other studies in the literature (see for example McCandless *et al.* 1972) the raw correlation between standardized test scores and teachers' marks is modest (systematically below 20 per cent). One could think that this result follows from the confounding effect of idiosyncrasies of teachers in reporting subjective measurements of achievement. This interpretation views standardized PISA tests as the "true" measurement, and teachers' marks as an error-ridden measurement of the same quantity.

Tab. 5 *Average teachers' marks in mathematics by tertiles of the standardized test score distribution.*

Remedial exam

| | Student cohort and School track | | | |
| | 2006/07 | | 2007/08 | |
Tertiles of test score	*Academic*	*Technical Vocational*	*Academic*	*Technical vocational*
Bottom (1st tertile)	6.885	6.167	6.676	6.297
Middle (2nd tertile)	7.156	6.500	7.156	6.712
Top (3rd tertile)	7.354	6.644	7.295	6.794
Corr Pisa score and Mark	0.141	0.034	0.227	0.172

Educational debt

| | Student cohort and School track | | | |
| | 2006/07 | | 2007/08 | |
Tertiles of test score	*Academic*	*Technical Vocational*	*Academic*	*Technical vocational*
Bottom	6.371	6.297	6.531	6.295
Middle	6.952	6.500	6.791	6.388
Top	6.769	6.747	6.800	6.700
Corr test score and Mark	0.197	0.185	0.167	0.147

Tab. 6 *Teachers effects on marks in mathematics by cohort and type of school (academic vis-à-vis technical/vocational schools).*

Treatment schools (remedial exam)

| | Student cohort and school type | | | |
| | 2006/07 | | 2007/08 | |
Tertiles of the Pisa score	*Academic*	*Technical Vocational*	*Academic*	*Technical vocational*
Bottom	0.548	0.111	0.117	0.008
Middle	0.229	0.060	0.593	0.035
Top	0.031	0.042	0.176	0.103

Control schools (educational debt)

| | Student cohort and school type | | | |
| | 2006/07 | | 2007/08 | |
Tertiles of the Pisa score	*Academic*	*Technical Vocational*	*Academic*	*Technical vocational*
Bottom	0.796	0.068	0.372	0.127
Middle	0.878	0.207	0.760	0.000
Top	0.068	0.379	0.751	0.083

Note: *Figures reported are p-values for the joint significance of class dummies estimated from separate regressions by cohort and type of school. For the cohort enrolled in 2006/07 (older cohort), results were obtained by regressing teachers' mark in mathematics at the end of the second year on (i) PISA scores (taken at the beginning of the following year), (ii) a dummy for students who passed at the end of the second year, and (iii) class dummies. For the cohort enrolled in 2007/08 (younger cohort), results were obtained by regressing teachers' mark in mathematics at the end of the first year on (i) PISA scores (taken at the beginning of the following year), (ii) a dummy for students who passed at the end of the first year, and (iii) class dummies.*

We investigated further this issue by regressing teachers' marks on PISA test scores and class dummies, the latter being included to model explicitly the presence of

teachers' fixed effects.[12] The regressions were run separately by tertiles of the distribution of marks across students. If marks solely reflect achievement, class dummies should jointly be insignificant in the regression once the PISA score is controlled for. In other words, class membership should not be a good predictor of teachers' marks once the "true achievement" is taken into account.

5.2   Effect of remedial exams

Tables 7 and 8 present the controlled differences in the test scores between treated and control schools. The two tables report estimated coefficients for the average treatment effect on the treated schools (ATT), which is the average gain from treatment for the treated schools (see Section 4.2). The main underlying assumption is that, in the absence of any remedial exam intervention, the average change in achievement would have been the same for treated and comparison schools.

Table 7 illustrates ATT estimates based on propensity score matching and obtained by comparing the test score of each treated student to a weighted average of outcomes of all control students, weights being lower for students with values of the propensity score that are far apart. The propensity score is constructed for all students using predicted values from a logistic regression of the treatment status dummy on student cohort-specific covariates. Four sets of covariates are introduced in the logistic regression: (i) *student socio-demographics* (gender, age, foreign student, cohabitation with mother and father, proximity to school) (ii) *socio-economic background* (father's age, father's education level, mother's age, mother's education level, housewife mother, unemployed mother or father, (iii) *household wealth and social-status* (occupational stratification scores, material deprivation index) (iv) *student's career*: school type (vocational or academic), year attended at interview (2nd or 3rd), final grade at lower secondary leaving certificate (8th grade), grade at the end of the first semester (1st year at secondary school) in Italian literature, maths and root-specific subjects, grade at the end of the first semester (2nd year at secondary school) in Italian literature, maths and root-specific subjects; dummy for passed/not passed at the end of the 1st year at secondary school; dummy for passed/not passed at the end of the 2nd year at secondary school. Weights are obtained using a Gaussian kernel as explained in Section 4.2 – see (4). The confidence intervals are obtained through a bootstrap procedure that makes use of 500 replications and clustering outcomes at the class level.

Table 8, which largely confirms results from the propensity score procedure, reports ATT coefficients based on (3) (that is a fully interacted linear regression model in which the treatment dummy is interacted with all covariates considered). Coefficient estimates are obtained from regressions of our assessment test scores on the same sets of

---

[12] The specification also included a dummy for students who passed at the end of the 2007/08 school year to capture the likely effects of additional schooling undertaken by students who were not admitted to the following year in the summer of 2008, and that would possibly affect their performance at the PISA test.

covariates used in the propensity score procedure and, in addition, a full set of interactions between the treatment dummy and each covariate. It is worth noting that the set of covariates used in the regressions is cohort-specific, reflecting the availability of pre-policy information. Thus, for the student cohort enrolled in 2006/07 (older cohort), we used teachers' marks at the first semester of the first year at school as well as the final grade at the lower secondary leaving certificate; while for the student cohort enrolled in 2007/08 (younger cohort), we could use the final grade at the lower secondary leaving certificate only. As in the propensity score strategy, the confidence intervals result from a bootstrap procedure, using 500 replications and clustering outcomes at the class level.

Reported outcomes indicate that the only statistically significant difference between treated and control students is in the discipline area of science and only for the older cohort of students attending academic tracks. Such students exhibit levels of scientific literacy higher than that of their peers who still face the *'educational debt'*. In substantive terms, the magnitude of such gain (32 score points on the PISA scientific scale) is a relatively large difference in student performance and can be interpreted as an advantage of about a school year ahead in terms of achieved proficiency.[13]

Tab. 7 *Propensity score matching estimates of the effects of remedial exams on student achievement by cohort and type of school (academic vis-à-vis technical/vocational schools).*

	Student cohort 2006/07			Student cohort 2007/08		
Academic schools	ATT	se	[95%conf int.]	ATT	se	[95%conf int.]
Reading	10.41	16.64	-9.33	52.77	11.20	16.81	-24.23	42.30
Mathematics	7.08	17.17	-22.65	45.41	13.64	16.16	-18.57	43.44
Science	**32.03**	12.18	9.28	58.89	15.30	12.88	-12.64	37.04
Technical/Vocational schools	ATT	se	[95%conf int.]	ATT	se	[95%conf int.]
Reading	-20.81	20.19	-58.81	26.85	-29.80	19.78	-63.32	12.28
Mathematics	-5.06	18.95	-32.24	33.44	-3.90	27.10	-60.62	52.20
Science	-2.43	18.69	-33.65	42.38	-25.40	23.45	-74.80	16.70

Tab. 8 *Fully interacted linear regression estimates of the effects of remedial exams on student achievement by cohort and type of school (academic vis-à-vis   technical/vocational schools).*

	Student cohort 2006/07			Student cohort 2007/08		
Academic schools	ATT	se	[95%conf int.]	ATT	se	[95%conf int.]
Reading	11.97	20.93	-21.53	51.10	6.84	17.72	-25.67	44.26
Mathematics	9.00	24.24	-23.14	49.11	9.89	14.30	-22.75	36.62
Science	**23.02**	25.51	4.06	57.35	16.97	9.67	-4.49	32.60
Technical/Vocational schools	ATT	se	[95%conf int.]	ATT	se	[95%conf int.]
Reading	-21.61	13.56	-45.85	8.36	-25.82	17.19	-61.85	6.11
Mathematics	0.08	14.82	-28.33	29.47	-11.77	20.03	-46.82	28.87
Science	-2.31	13.45	-26.96	23.67	-24.71	17.97	-57.97	14.94

Note: *Results were obtained by running a regression of test scores on a dummy for treatment /control areas, individual and school characteristics, and a full set of interactions.*

---

[13] One school year corresponds to an average of 38 score points on the PISA scientific scale (OECD 2007b, p.55)

Although the only statistically significant difference is observed among a definite student group, what we find striking and noteworthy is the fact that overall results reveal a clear pattern of positive effects for students attending academic tracks and negative effects for students on technical tracks. Both tables 7 and 8 show that students attending academic secondary schools under the remedial exam procedure perform better in *all* tested domains than students not subject to it. On the opposite, students of technical schools undergoing the remedial exam procedure show levels of reading, mathematical and scientific literacy lower than those of their counterparts who do not face the new progression system. Leaving statistical significance aside, the results clearly point to the existence of systematic disparities in achievement levels between treated and control schools. Such disparities are systematically positive among students from academically oriented schools and systematically negative among students from technical and vocational schools.

Nonetheless, it is worth commenting on the impact of remedial exams on scientific literacy. Why has the intervention a stronger impact on scientific literacy than on mathematics and reading literacy? Why is such impact more remarkable for students in the older cohort? Typically, in Italian upper secondary schools scientific subjects are regarded as marginal disciplines both by students and teachers. However, under the new progression rules, failing even in one single subject can lead to the remedial exam. As a consequence, students are compelled to reach adequate achievement standards in *all* subjects - including science – if they intend to avoid grade retention. This new rule appears to have positively motivated students to do particularly well in assessment areas previously considered as minor. The magnitude of the effect for the older cohort is noteworthy and may well be explained by the fact that the older cohort experienced their first school year (2006/07) under the old 'debt' regime unlike the younger cohort. Thus, to remediate the accumulated lack in achievement in subjects previously neglected, they may have put an extra effort to do well. Whereas, the younger cohort never experienced the old progression system and therefore did not feel the urge or pressure to do better.

We have experimented with a variety of specifications, variables and estimation techniques and findings are found to be robust. Particularly, we did sensitivity analysis to test the impact of teachers' marks on the regression results. For example, for the older student cohort (2006/07) we experimented with two different sets of control variables: (i) "only pre-reform controls" and (ii) "also post-reform controls". The first set consists of the control variables introduced in the previous models (*student socio-demographic; socio-economic background*; *household wealth and social-status)* and controls for the student's career that refer to the pre-reform period: final grade the lower secondary leaving certificate and teachers' marks during the first year at upper secondary school. The second set also controls for covariates that refer to the *post-reform* period: teachers' marks during the first semester of the second year, that is the closest available marks to the introduction of the reform. Such marks are potentially affected by the reform, and should be treated as outcome variables rather than as controls. Nevertheless, we did so because, by design, for the younger cohort (2007/08) we have mainly administrative information from schools referred to post-reform periods. So, for the younger cohort we can control for family background information and final grade at lower secondary leaving certificate ("only *pre-reform* controls"), or family background information, final grade at lower secondary

leaving certificate and teachers' marks during the first year at school ("also *post-reform* controls"). Sensitivity analyses are available in the Appendix (tables A1, A2, A3 and A4). Results from this exercise are consistent with those from the analyses presented in this section.

6.   Discussion and conclusions

Remedial programmes for compulsory secondary education have received relatively scant attention. We have presented evidence that such programmes have mixed effect on student achievement. Our findings consistently depict positive point estimates for the effect of remedial exams on the performance of students enrolled in academic tracks, and negative point estimates for students in technical and vocational oriented schools. This result holds across different school years and applies to all discipline areas considered (i.e. reading, mathematics and science). Despite the regularity of this pattern, only the positive effect on test scores that refer to scientific competencies in academic schools is found significant, and amounts to nearly one additional year of school in terms of achieved proficiency.

In interpreting this interesting pattern, it is crucial to bear in mind three major aspects of any educational system: (i) the influence of parental social background on children's educational outcomes, (ii) the role of risk aversion in educational choices, and (iii) the impact of school tracking on student achievement. In all industrialised countries the social background of origin has been found to greatly influence children's educational attainment. As in other advanced countries, a large proportion of Italian students attending academic tracks consists of children from advantaged social backgrounds, while students in technical and vocational schools are, on average, more likely to come from less privileged backgrounds. Secondly, education is an investment. Students choose to put effort and time in school to balance its social return with its opportunity cost. Thus, when students (and their families) are faced with risky choices about their educational career, they exhibit risk aversion that leads them to minimize the risk of failing. Thirdly, most secondary school systems maintain the practice of educational tracking, that is placing children into different tracks that prepare some students for college or university and others for vocational skills that do not lead to college or university (Shavit & Muller 2000). Arguments against secondary school tracking suggest that this practice creates homogeneous classes according to ability and social background, and reduces the positive spillover effect known as peer effect (Zimmer 2003; Hanushek & Wößmann 2005).

By making the important distinction between *primary* and *secondary* effects, Boudon (1974) has demonstrated that parents can affect their children's educational attainment in at least two ways. Disparities in educational transitions between children of different social backgrounds can be due to the different impact parents exert on their children's cognitive ability (*primary effects*, e.g. high status parents are more likely to

contribute to create a stimulating cultural environment for their offspring). Yet, families can also influence their children performance by making choices within the educational system, *given* the level of cognitive ability (*secondary effects*). Hence, children of advantaged social classes typically show more positive attitudes towards school and learning, and feel more confident in their school performance (Burnhill *et al.*, 1990; Paterson 1991). Particularly, previous studies have demonstrated that students from higher socio-economic status not only show better performance on the reading, mathematical and scientific literacy (OECD 2001), but also exhibit higher levels of student engagement, that is the extent to which students identify with and value schooling outcomes, and participate in academic and non-academic school activities (Wilms 2003). Our evidence on students of academic oriented schools demonstrates that when faced with the risk of grade retention, they are positively motivated to engage in further studying in order to avoid failure. Whereas, in the lack of any punishment (e.g. grade retention) or binding assessment of their performance (e.g. remedial exam), even students from privileged social backgrounds tend to minimise their effort to pass to the next school grade. Following up on Boudon (1974), Breen and Goldthorpe (1997) argue that children aim to acquire a social position at least as advantageous as that of their parents. The relative risk aversion, namely the concern with downward social mobility, therefore varies between social classes and can affect schooling ambitions differently. Students will make choices on their educational career and will remain in school until their aspirations are fulfilled. We showed that the threat of grade retention appears not to enhance study motivation of pupils with lower aspirations but rather to reduce their commitment to it. Students in those technical schools where the 'educational debt procedure' was still ongoing performed, on average, better than those subject to the new progression system. Clearly, the relative risk aversion is higher for children of lower social origins. Thereby, they are more reluctant to accept additional commitment to school (e.g. extra study time in preparation of some remedial exam) with an uncertain payoff rather than minimising their effort to obtain a more certain, and possibly lower, payoff.

Peer pressure as well as teachers' expectations are also likely to have played a crucial role in shaping students' educational responses to the new progression system. Not only students perform at a higher level if their peers are high achievers (Evans *et al.* 1992), but peers can also act as a buffer by legitimising deviant behaviour. As well known, grade retention and dropping-out of school are not uncommon events in vocational and technical schools. The social stigma attached to such events is therefore likely to be smaller for children attending schools in which a non-negligible share of the population experience them. Previous research has also demonstrated that whether they are high or low, teachers' expectations have a significant impact on overall student achievement. High teachers' expectations has been found to be a consistent positive predictor of students' goals and interests, while negative feedback is the most consistent negative predictor of academic performance (Wentzel 2002). More generally, how teachers manage classroom culture (e.g. creating a supportive environment for low-achieving students) has a strong impact on student motivation and passing rates (Roderick and Engel 2001; Akerlof & Cranton 2003). Given that teachers' expectations have been found to vary depending on socio-economic status and ethnicity of students (Gollub 1978; van den Bergh *et al.* 2010),

it seems reasonable to hypothesise that expectations of teachers in academic schools are higher than those of teachers in schools such as technical/vocational schools where, conversely, students typically show a higher susceptibility to negative teacher expectancy effects.

As it stands now, the intervention appears to reinforce, at least in the short run, preexisting inequalities in achievement among students from different school tracks. In line with previous research on educational tracking, our results reveal clear differentials in the students' responses to the introduction of the remedial exam practice. Specifically, the threat of punishment appears to be effective in decreasing undesired behaviour only among students in academic tracks and who presumably possess more positive attitudes towards school learning; while in the case of students in technical/vocational tracks the threat of punishment does not produce the expected outcome. Therefore, punishment does not directly imply a decreasing of undesired behaviour for all students but it is likely to interact with a multiplicity of factors.

These findings have important policy implications. In order to help close this 'achievement gap', vocational students and, more generally, students from socially disadvantaged backgrounds are likely to benefit more from reinforcement of positive attitudes towards learning as well as reinforcement of positive values about their schooling outcomes rather than punishment practices. The main drawback of the remedial exam practice is that it appears to undermine effort and engagement, especially of struggling students. As repeatedly shown, study effort is responsive to positive incentives rather than punishment. For instance, a growing number of both developing and developed countries are providing cash transfer to students from disadvantaged backgrounds to help them stay in school[14]. In most cases the transfer's impact on education participation rates is substantial (Bourgignon *et al.* 2003; Berhman *et al.* 2005; Dearden *et al.* 2009). Besides, merit-based scholarships have been recently found to raise not only school attendance but also test scores and boost classroom effort (Kremer *et al.* 2009). In Italy, the introduction of the minimum income has remarkably decreased the drop-out rates by about 50 per cent in towns like Naples and Catania among children from disadvantaged backgrounds (Saraceno 2002). In addition, recent comparative studies reveal that students perform significantly better in countries where large shares of schools use accountability measures to make decisions about students' retention or promotion (Wößmann *et al.* 2007). Nonetheless, our results suggest that - in order for school tracking to not further enhance educational inequality - accountability policies would be more effective in increasing appropriate behaviour if they were targeted and designed for specific student populations.

---

[14] The first cash transfer programmes conditional on children's school attendance were implemented in Brazil in 1995 (*Bolsa Família*) and in Mexico in 1997 (*Progresa*). Low-income families typically receive a monthly allowance conditional upon sending their children to school and perform other activities, such as ensuring their children receive regular health check-ups. More recent examples are the Opportunity NYC programme in the US and the Education Maintenance Allowance (EMA) in the UK (Dearden *et al.*2009).

As a matter of fact, implementing the very same policy for dissimilar individuals has never proven to enhance social equality. Such policies ultimately ignore the complexity of children's learning system, the multidimensional nature of the problem and the determinants of intrinsic motivation. A simple 'stimulus-response' approach to the improvement of student performance is not sufficient because does not take into account substantial differences between students in terms of (i) learning mechanisms (abilities, motivation), (ii) peer effects, and expectations of (iii) teachers and (iv) parents. Any set of policies that put responsibility only on students will fall short in the end. Future research is therefore needed to shed light on the mechanisms underlying the differential effects of remedial education. Our analysis points to the complexity of factors involved in this process. However, the interconnectedness among the determinants of this process remains largely unexplored. Further research is required to assess whether the impact of the remedial programme applies not only to short-term achievement gains/losses but also to long-term learning.

References

Akerlof, G. & Kranton, R. (2002) Identity and schooling: Some lessons for the economics of education. *Journal of Economic Literature,* 40(4)**,** 1167-1201.

Alfred, R. & Lum, G. (1988) Remedial programme policies, student demographic characteristics, and academic achievement in community colleges. *Community/Junior College Quarterly of Research and Practice,* 12(2)**,** 107-20.

Allensworth, E. (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.

Angrist, J. & Lavy, V. (1999) Using Maimonides' rule to estimate the effect of class size on scholastic achievement*. *Quarterly journal of economics,* 114(2)**,** 533-575.

Behrman, J., Sengupta, P. & Todd, P. (2005) Progressing through PROGRESA: An impact assessment of a school subsidy experiment in rural Mexico. *Economic Development and Cultural Change,* 54(1)**,** 237.

Benabou, R. & Tirole, J. (2003) Intrinsic and extrinsic motivation. *Review of Economic Studies,* 70(3)**,** 489-520.

Benadusi, L. & Consoli, F. (Eds.) (2004) *La governance della scuola. Istituzioni e soggetti alla prova dell'autonomia,* (Bologna, Il mulino).

Bettinger, E. & Long, B. (2007) Institutional responses to reduce inequalities in college outcomes: Remedial and developmental courses in higher education, in: S. Dickert-Conlin & R. Rubenstein (Eds) *Economic inequality and higher education: Access, persistence, and success.* 69-100.

Black, D. & Smith, J. (2006) Estimating the returns to college quality with multiple proxies for quality. *Journal of Labor Economics,* 24(3)**,** 701-728.

Bottani, N. (1986) *La ricreazione è finita,* (Bologna, Il mulino).

Boudon, R. (1974) *Education, opportunity, and social inequality: Changing prospects in western society,* (New York, Wiley).

Bourguignon, F., Ferreira, F. & Leite, P. (2003) Conditional cash transfers, schooling, and child labor: Micro-simulating Brazil's Bolsa Escola programme. *The World Bank Economic Review,* 17(2)**,** 229.

Boylan, H. R., Bonham, B. S. & White, S. R. (1999) Developmental and remedial education in postsecondary education. *New Directions for Higher Education,* 1999(108)**,** 87-101.

Breen, R. & Goldthorpe, J. (1997) Explaining educational differentials: Towards a formal rational action theory. *Rationality and Society,* 9(3)**,** 275.

Bronstein, P., Ginsburg, G. & Herrera, I. (2005) Parental predictors of motivational orientation in early adolescence: A longitudinal study. *Journal of Youth and Adolescence,* 34(6)**,** 559-575.

Burnhill, P., Garner, C. & Mcpherson, A. (1990) Parental education, social class and entry to higher education 1976-86. *Journal of the Royal Statistical Society. Series A (Statistics in Society),* 153(2)**,** 233-248.

Calcagno, J. & Long, B. (2008) The impact of postsecondary remediation using a regression discontinuity approach: Addressing endogenous sorting and noncompliance. Cambridge Mass., USA, National Bureau of Economic Research.

Cobalti, A. & Schizzerotto, A. (1993) Inequality of educational opportunity in italy, in: Y. Shavit & H. P. Blossfeld (Eds) *Persistent inequality. Changing educational attainment in thirteen countries.* Boulder, Westview Press), 155–176.

Collier, T. & Millimet, D. (2009) Institutional arrangements in educational systems and student achievement: A cross-national analysis. *Empirical Economics,* 37(2)**,** 329-381.

De Lillo, A. & Schizzerotto, A. (1985) *La valutazione sociale delle occupazioni: Una scala di stratificazione occupazionale per l'Italia contemporanea,* (Bologna, Il Mulino).

Dearden, L., Emmerson, C., Frayne, C. & Meghir, C. (2009) Conditional cash transfers and school dropout rates. *Journal of Human Resources,* 44(4)**,** 827.

Deci, E., Koestner, R. & Ryan, R. (1999) A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological bulletin,* 125(627-668.

Dobbelsteen, S., Levin, J. & Oosterbeek, H. (2002) The causal effect of class size on scholastic achievement: Distinguishing the pure class size effect from the effect of changes in class composition. *Oxford Bulletin of Economics and Statistics,* 64(1)**,** 17-38.

Erikson, R., Goldthorpe, J., Jackson, M., Yaish, M. & Cox, D. (2005) On class differentials in educational attainment. *Proceedings of the National Academy of Sciences,* 102(27)**,** 9730.

Eurydice (2009) *Organisation of the education system in Italy 2008/09.* Report for the European Commission.

Evans, W., Oates, W. & Schwab, R. (1992) Measuring peer group effects: A study of teenage behavior. *The Journal of Political Economy,* 100(5)**,** 966-991.

Gasperoni, G. (1996) *Diplomati e istruiti. Rendimento scolastico e istruzione secondaria superiore,* (Bologna, Il mulino).

Goldthorpe, J. (1996) Class analysis and the reorientation of class theory: The case of persisting differentials in educational attainment. *British Journal of Sociology***,** 481-505.

Gollub, W. & Sloan, E. (1978) Teacher expectations and race and socioeconomic status. *Urban education,* 13(1)**,** 95.

Greenwald, R., Hedges, L. & Laine, R. (1996) The effect of school resources on student achievement. *Review of Educational Research,* 66(3)**,** 361.

Grubb, W. (2001) From black box to Pandora's box: Evaluating remedial/developmental education (CRCC brief no. 11). *New York: Community College Research Center, Teachers College, Columbia University.*

Hamilton, L. (2003) Assessment as a policy tool. *Review of Research in Education,* 27(1)**,** 25–68.

Hanushek, E. (1997) Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis,* 19(2)**,** 141.

Hanushek, E., Kain, J., Markman, J. & Rivkin, S. (2003) Does peer ability affect student achievement? *Journal of Applied Econometrics,* 18(5)**,** 527-544.

Hanushek, E., Wößmann, L. & Str, P. (2005) Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *NBER Working Paper.*

Heckman, J., Lalonde, R. & Smith, J. (1999) The economics and econometrics of active labor market programmes. *Handbooks in Economics,* 5(3 PART A)**,** 1865-2095.

Heckman, J. & Robb Jr, R. (1985) Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics,* 30(1-2)**,** 239-267.

Hoxby, C. (2000) The effects of class size on student achievement: New evidence from population variation. *Quarterly journal of economics,* 115(4)**,** 1239-1285.

Ishida, H. & Ridge, J. (1995) Class origin, class destination, and education: A cross-national study of ten industrial nations. *The American Journal of Sociology,* 101(1)**,** 145-193.

Jackson, M., Erikson, R., Goldthorpe, J. & Yaish, M. (2007) Primary and secondary effects in class differentials in educational attainment: The transition to a-level courses in England and wales. *Acta Sociologica,* 50(3)**,** 211.

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

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

Kruglanski, A. (1978) Issues in cognitive social psychology. *The Hidden Cost of Reward: New Perspectives on the Psychology of Human Motivation.*

Lavy, V. (2003) *Paying for performance: The effect of individual financial incentives on teachers' productivity and students' scholastic outcomes.* Report for Working Paper, The Hebrew University of Jerusalem and CEPR.

Lavy, V. & Schlosser, A. (2005) Targeted remedial education for underperforming teenagers: Costs and benefits. *Journal of Labor Economics,* 23(4)**,** 839-874.

Linn, R., Baker, E. & Dunbar, S. (1991) Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher,* 20(8)**,** 15.

Masters, G. (1982) A Rasch model for partial credit scoring. *Psychometrika,* 47(2)**,** 149-174.

Mccandless, B., Roberts, A. & Starnes, T. (1972) Teachers' marks, achievement test scores, and aptitude relations with respect to social class, race, and sex. *Journal of Educational Psychology,* 63(2)**,** 153-159.

Merisotis, J. & Phipps, R. (2000) Remedial education in colleges and universities: What's really going on? *Review of Higher Education,* 24(1)**,** 67-86.

Mistry, R., White, E., Benner, A. & Huynh, V. (2009) A longitudinal study of the simultaneous influence of mothers' and teachers' educational expectations on low-income youth's academic achievement. *Journal of Youth and Adolescence,* 38(6)**,** 826-838.

Nolan, B., Gannon, B., Layte, R., Watson, D., Whelan, C. & Williams, J. (2002) Monitoring poverty trends in Ireland: Results from the 2000 living in Ireland survey. *Research Series.*

OECD (2001) *Knowledge and skills for life. First results from PISA 2000,* (Paris, OECD).

OECD (2004) *Learning for tomorrow's world. First results from PISA 2003,* (Paris, OECD).

OECD (2007a) *Evidence in education: Linking research and policy,* (Paris, OECD).

OECD (2007b) *Pisa 2006 science competencies for tomorrows world: Volume 1 analysis,* (Paris, OECD).

Parsad, B., Lewis, L. & Greene, B. (2003) Remedial education at degree-granting postsecondary institutions in fall 2000. *Institute for Educational Sciences. US Department of Education. Retrieved February,* 16(2007.

Paterson, L. (1991) Socio-economic status and educational attainment: A multi-dimensional and multi-level study. *Evaluation & Research in Education,* 5(3)**,** 97-121.

Robin, S. & Sprietsma, M. (2003) Characteristics of teaching institutions and student's performance: New empirical evidence from OECD data. *IRES Discussion Papers,* 28

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

Rosenbaum, P. & Rubin, D. (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika,* 70(1)**,** 41.

Rubin, D. (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology,* 66(5)**,** 688-701.

Sacerdote, B. (2001) Peer effects with random assignment: Results for Dartmouth roommates. *Quarterly journal of Economics,* 116(2)**,** 681-704.

Saraceno, C. (Ed.) (2002) *Rapporto sulle politiche contro la povertà e l'esclusione sociale: 1997-2001* (Roma, Carocci).

Shavit, Y. & Müller, W. (2000) Vocational secondary education, tracking, and social stratification. *Handbook of the sociology of education***,** 437-452.

Skinner, B. (1938) *The behavior of organisms: An experimental analysis,* (New York, Appleton-Century-Crofts).

Spera, C., Wentzel, K. & Matto, H. (2009) Parental aspirations for their children's educational attainment: Relations to ethnicity, parental education, children's academic performance, and parental perceptions of school climate. *Journal of Youth and Adolescence,* 38(8)**,** 1140-1152.

Staddon, J. (1983) *Adaptive learning and behavior,* (Cambridge: Cambridge University Press).

Todd, P. & Wolpin, K. (2003) On the specification and estimation of the production function for cognitive achievement. *The Economic Journal,* 113(485)**,** 3-33.

Van Den Bergh, L., Denessen, E., Hornstra, L., Voeten, M. & Holland, R. W. (2010) The implicit prejudiced attitudes of teachers: Relations to teacher expectations and the ethnic achievement gap. *American Educational Research Journal***,** 0002831209353594.

Wentzel, K. (2002) Are effective teachers like good parents? Teaching styles and student adjustment in early adolescence. *Child Development,* 73(1)**,** 287-301.

Whelan, C., Layte, R. & Maitre, B. (2002) Multiple deprivation and persistent poverty in the European Union. *Journal of European Social Policy,* 12(2)**,** 91.

Willms, J. (2003) *Student engagement at school: A sense of belonging and participation. Results from PISA 2000,* (Paris, OECD).

Wößmann, L. (2003) Schooling resources, educational institutions and student performance: The international evidence. *Oxford Bulletin of Economics and Statistics,* 65(2)**,** 117-170.

Wößmann, L., Lüdemann, E., Schütz, G. & West, M. (2007) School accountability, autonomy, choice, and the level of student achievement: International evidence from PISA 2003. *OECD Education Working Paper: EDU/WKP(2007)8*.

Zimmer, R. (2003) A new twist in the educational tracking debate. *Economics of Education Review,* 22(3)**,** 307-315.

Appendix

The aim of this appendix is to discuss the effects of adjusting for teachers' marks during the first year at secondary school for the younger cohort of students considered in the analysis. As it is clear from figure 1, performance at school for this cohort in the first year is potentially affected by the reform that was introduced prior to enrolment. Because of this, the adjustment made to control for compositional differences in the two groups of school involves concomitant *variables* that should themselves be considered as "outcomes".

This setup closely resembles that encountered in other studies on cognitive achievement, where for example senior test scores are adjusted using sophomore test scores which are themselves potentially affected by the status of individuals with respect to the "intervention" (e.g. Coleman, Hoffer and Kilgore, 1982). In this instance, sophomore test scores are used to proxy unobserved achievement *before* high school, just as we proxy achievement at the beginning of upper secondary school with administrative data on teachers' marks at the end of the first semester of the first year.

The consequences for identification of adjusting for covariates affected by the reform are discussed by Rosenbaum (1984). The general problem can be easily put across by considering a situation in which the causal effect on achievement that follows from membership to the $D = 1$ group would be recovered by conditioning on the variables (X, U). The former is a set of observable variables, whereas the latter set refers to achievement outcomes pre-determined with respect to the reform and *not* observable in the data. Let (S_0, S_1) be the potential outcomes of the *concomitant* variables used to proxy the unobserved component U, whose interpretation coincides with that for (Y_0, Y_1). In our setup, they represent the closest measurement in time to the covariate U.

We will focus on the situation in which (i) controlling for X alone fails to retrieve the causal parameter of interest, and (ii) differences in the composition of students are accounted for by aligning the $D = 0$ and $D = 1$ groups with respect to (X, S_0) and (X, S_1), respectively. Throughout our discussion we will make the two following assumptions.

Assumption 1: *the conditional independence assumption holds for the pair (Y_0, S_0) with respect to the variables (X, U), of which only the former is observable.*

According to this assumption there is (for d = 0,1):

$$E[Y_0|D = d, X = x, U = u] = E[Y_0|X = x, U = u],$$

$$E[S_0|D = d, X = x, U = u] = E[S_0|X = x, U = u].$$

Assumption 2: *the post treatment concomitant measurements (S_0, S_1) are a surrogate for the unobserved pre-treatment variable U:*

$$E[Y_0|S_0 = s, X = x, U = u] = E[Y_0|S_0 = s, X = x].$$

$$E[Y_1|S_1 = s, X = x, U = u] = E[Y_1|S_1 = s, X = x].$$

Assumption 1 simply extends the ignorability condition to all potential outcomes. Assumption 2 embodies the idea that the conditioning on either S_0 or S_1 is sufficient for the information relevant to the outcome of interest contained in the unobserved component U. As pointed out by Rosenbaum (1984), this appears to be a plausible assumption when

28

U and its surrogates – as it is in our case – represent consecutive measures of achievement over time.

Aligning the two groups of schools with respect to (X, S_0) and (X, S_1) will in general lead to incorrect inference on the causal parameter of interest. To see this, start from Assumption 1 and write:

$$E[Y_0|D = 1] = \int E[Y_0|X = x, U = u]f(u|D = 1, X = x)f(x|D = 1)dudx.$$

Then use Assumptions 1 and 2 to write:

$$E[Y_0|X = x, U = u] = \int E[Y_0|D = 0, S_0 = s, X = x]\, f_{S_0}(s|X = x, U = u)ds,$$

so that the last two expressions combined yield:

$$E[Y_0|D = 1] = \int E[Y_0|D = 0, S_0 = s, X = x]\, f_{S_0 X}(s, x|D = 1)dxds.$$

The identification problem is clearly spelled out by considering the last expression. The term $E[Y_0|D = 0, S_0 = s, X = x]$ can be retrieved from information on (Y_0, S_0, X) in the $D = 0$ group, which is observable in the data. The distribution of (S_0, X) for the $D = 1$ group is however *not* observable by design, as realization of D would reveal S_0 only for $D = 0$ schools. If we ignore, or don't perceive, that concomitant outcomes are potentially affected but the reform, we would implicitly make the following assumption.

Assumption 3: *the concomitant measurements (S_0, S_1) are not affected by the treatment:*

$$f_{S_0}(s|D = 1, X = x) = f_{S_1}(s|D = 1, X = x).$$

Under Assumption 3, estimates of the effect could for example be obtained by matching students across schools with respect to X and *observed* values of S, where $S = S_0$ for $D = 0$ schools and $S = S_1$ for $D = 1$ schools. Assumptions 1, 2 and 3 are thus key to identification (see the discussion in Rosenbaum, 1984, Section 3.6). Note that Assumption 3 amounts to saying that, conditional on X, the distribution of teachers' marks is stable under the two policy regimes in schools outside the administrative border of the Trento province.

In the context dealt with in this paper, Assumption 3 appears to be the most questionable. Since we can write (note that a support condition is required to ensure quantities in the equation are meaningfully defined):

$$E[Y_0|D = 1] = \int E[Y_0|D = 0, S_0 = s, X = x]\, \frac{f_{S_0}(s|D=1,X=x)}{f_{S_1}(s|D=1,X=x)} f_{S_1 X}(s, x|D = 1)dxds,$$

it follows that treatment effects can be retrieved only by making assumptions on the distribution of (S_0, S_1) in $D = 1$ schools. For example, if one were willing to assume (for all x's):

$$f_{S_0}(s|D = 1, X = x) \leq f_{S_1}(s|D = 1, X = x),$$

there would be:

$$E[Y_0|D = 1] \leq \int E[Y_0|D = 0, S_0 = s, X = x]\, f_{S_1 X}(s, x|D = 1)dxds,$$

so that point estimates obtained ignoring the problem could be used to bound the true effect.

We instead take a different route, and look at the distributions of teachers' marks across the two cohorts of students. Start by comparing the distribution of marks in the first year for schools with D=0:

$$f_{S_0}(s|D = 0, X = x)$$

for the two cohorts of students. All students enrolled in the first year should be considered, regardless of whether they went through the PISA test. Run a regression of S on X and a cohort dummy for the D=0 sample. This is a test of the stability of the distribution across cohorts.

Tab.A1 *Sensitivity analysis: effects of remedial exams on student achievement by type of school (general vis-à-vis technical/vocational schools) for the cohort enrolled in 2006/07.*

General education schools	ATT	se	[95%conf int.]	
Reading	8.36	14.35	-23.69	36.61
Mathematics	17.71	18.93	-16.71	47.56
Science	22.66	12.41	-0.59	48.71
Technical/Vocational schools	ATT	se	[95%conf int.]	
Reading	-23.52	12.72	-46.84	5.26
Mathematics	3.21	14.27	-26.92	29.01
Science	-3.68	12.34	-26.43	22.47

Note: Results were obtained from regressions of PISA measurements on a dummy for schools implementing remedial exams, household and student demographics, teachers' marks and a full set of interactions with the treatment dummy (see Section 4.2 for more details about the estimation procedure). Covariates include (i) household and student demographics, and (ii) final grade at the lower secondary level. Figures in the table refer to the average treatment effect on the treated, that was defined in Section 4.2. The confidence intervals reported were obtained through a bootstrap procedure, using 500 replications and clustering outcomes at the class level.

Tab.A2 *Sensitivity analysis: propensity score matching estimates of the effects of remedial exams on student achievement by type of school (general vis-à-vis technical/vocational schools) for the cohort enrolled in 2006/07.*

General education schools	ATT	se	[95%conf int.]	
Reading	0.39	17.13	-20.71	49.88
Mathematics	-0.57	13.10	-20.63	31.81
Science	**21.94**	**10.90**	**6.34**	**49.66**
Technical/Vocational schools	ATT	se	[95%conf int.]	
Reading	-25.10	17.05	-52.20	18.18
Mathematics	5.27	17.30	-27.50	41.85
Science	-4.09	16.77	-33.62	33.03

Note: Results were obtained by comparing the outcome of each treated student to a weighted average of outcomes of all control students, weights being lower for students with values of the propensity score that are far apart. The propensity score was constructed for all students using predicted values from a logistic regression of the treatment status dummy. Weights were obtained using a Gaussian kernel as explained in Section 4.2. Figures in the table refer to the average treatment effect on the treated. The confidence intervals reported were obtained through a bootstrap procedure, using 500 replications and clustering outcomes at the class level.

Tab.A3 *Sensitivity analysis II: effects of remedial exams on student achievement by type of school (general vis-à-vis technical/vocational schools).*

	Cohort 2006/07				Cohort 2007/08			
General education schools	ATT	se	[95%conf int.]		ATT	se	[95%conf int.]	
Reading	2.04	24.03	-25.09	64.01	-1.21	22.60	-53.03	35.94
Mathematics	11.15	27.95	-26.68	58.19	5.92	16.87	-25.60	33.51
Science	21.03	28.10	-1.53	62.25	7.62	30.60	-25.51	31.49
Technical/Vocational schools	ATT	se	[95%conf int.]		ATT	se	[95%conf int.]	
Reading	-29.58	15.62	-58.07	5.58	-18.60	18.50	-53.16	18.67
Mathematics	-6.35	14.53	-37.36	20.56	-7.98	22.07	-49.82	36.07
Science	-7.37	15.22	-36.23	23.68	-11.62	18.78	-48.17	22.67

Note: Results were obtained from regressions of PISA measurements on a dummy for schools implementing remedial exams, household and student demographics, teachers' marks and a full set of interactions with the treatment dummy (see Section 4.2 for more details about the estimation procedure). The set of covariates used in the regressions was cohort-specific: for the cohort enrolled in 2006/07 (older cohort) we used (i) household and student demographics, (ii) teachers' marks at the first semester of the first year at school, and (iii) final grade at the lower secondary level (iv) teachers' marks at the first semester of the second year at school. For the cohort enrolled in 2007/08 (younger cohort) we used (i) household and student demographics, and (ii) final grade at the lower secondary level (iii) teachers' marks at the first semester of the first year at school. Figures in the table refer to the average treatment effect on the treated, that was defined in Section X. The confidence intervals reported were obtained through a bootstrap procedure, using 500 replications and clustering outcomes at the class level.

Tab. A4 *Sensitivity analysis II: propensity score matching estimates of the effects of remedial exams on student achievement by type of school (general vis-à-vis technical/vocational schools).*

	Cohort 2006/07				Cohort 2007/08			
General education schools	ATT	se	[95%conf int.]		ATT	se	[95%conf int.]	
Reading	28.42	16.40	-1.36	62.64	3.51	20.75	-43.08	36.72
Mathematics	18.06	19.87	-20.27	60.99	3.92	17.30	-24.08	41.15
Science	**34.64**	**16.82**	**8.17**	**77.45**	0.22	16.42	-28.38	35.46
Technical/Vocational schools	ATT	se	[95%conf int.]		ATT	se	[95%conf int.]	
Reading	-22.55	22.01	-60.12	23.88	-18.18	20.46	-57.81	19.37
Mathematics	-12.22	20.70	-45.14	38.14	-4.97	26.91	-64.66	44.88
Science	-3.26	19.91	-42.41	43.19	-16.30	24.12	-68.83	27.64

Note: Results were obtained by comparing the outcome of each treated student to a weighted average of outcomes of all control students, weights being lower for students with values of the propensity score that are far apart. The propensity score was constructed for all students using predicted values from a logistic regression of the treatment status dummy on cohort-specific demographics. Weights were obtained using a Gaussian kernel as explained in Section 4.2. Figures in the table refer to the average treatment effect on the treated. The confidence intervals reported were obtained through a bootstrap procedure, using 500 replications and clustering outcomes at the class level.