Threat of Grade Retention, Remedial Education and Student Achievement: Evidence from Upper Secondary Schools in Italy

Erich Battistin

Queen Mary University of London, CEPR, FBK-IRVAPP and IZA

Antonio Schizzerotto

University of Trento and FBK-IRVAPP

Abstract

We use a reform in Italy to investigate the effects on academic achievement of more stringent requirements for grade progression at secondary school. Geographic discontinuities in the implementation allow the comparison of similar students undergoing alternative progression rules. We find different short term effects across curricular tracks, with negative effects of the increased threat of grade retention for students in technical and vocational schools. In contrast with the effects on academic achievement, we find that schools reacted to the additional administrative burden and costs imposed by the reform by admitting more students to the next grade. We argue that the reform had a negative effect on motivation and engagement of the most struggling students, exacerbating inequalities.

1. Introduction

Concerns about education quality motivate the implementation of accountability policies to hold administrators, teachers and students responsible for academic achievement. The rationale for this strategy is that the promise of rewards or the threat of sanctions are needed to ensure change and set clear standards to promote educational change. The rise in accountability has increasingly led schools and teachers to rely on summer and remediation courses to help low-achieving students meet minimum standards. This paper assesses the effects on academic performance of a major remedial education reform in Italy.

Starting from the year 2007/08, students at secondary school lagging behind predefined standards of performance must attend remedial summer courses. Progression to the next grade is conditional on passing an exam before the new school year. This rule replaced a system in which the practice of social promotion was effectively at work. Before the reform students could be admitted to the next grade with an 'educational debt' (*debito formativo*) signalling lack of expected achievement, but with limited implications for progression across grades. Moreover, the reform gave to remedial education a more prominent role in the formative plan, making it mandatory during the year and in the summer.

The policy question addressed in this paper is whether mandating remedial summer courses for low-achieving students, and making admission to the next grade conditional on exams after the summer, makes a difference. In particular, we study the short-term effects on academic achievement, and how these are mediated by possible changes in school and teacher inputs. Our empirical strategy exploits the quasi-experimental variation resulting from a geographic discontinuity in the implementation of the reform. Unlike the rest of the country, schools located in a well-defined area of Northern Italy, the province of Trento, were exempted from adopting the new progression system. We use this setting to obtain counterfactual quantities and draw causal conclusions.

Remedial exams were introduced in Italy in 1923, and were abolished in secondary schools during the 1990s. Their reintroduction in 2007 resulted from political pressure after poor performance at PISA tests. Variability in scores across regions pictured a sharp North/South divide, with students in the South performing well below the OECD average. In addition, marked differences emerged across curricular tracks, even in the best performing areas. Advocates of the reform believed that the threat of grade retention was the most effective solution: pupils would study more intensively, as they instinctively fear failure. This view echoes the reinforcement theory of behaviourist psychology (Staddon, 2003). Reform opponents, instead, raised the concern that the threat of grade retention might undermine effort, motivation and engagement of struggling students, thus exacerbating inequalities.[1]

We evaluate the reform using survey and administrative data from complementary sources of information. Scores and socio-economic indicators come from a small scale survey for a sample of schools located either side of the administrative border of the province of Trento. This information is integrated with administrative data on retention rates for schools in the areas considered for the analysis. The analysis employs schools of the province of Trento as a control group for schools located the either side of the administrative border. We use census and administrative data to show that schools located in "treatment" and "control" areas share similar demographic structure and labour market outcomes before the reform.

Our findings can be summarized as follows. We document a large, negative effect on academic achievement (0.3 to 0.4 points of standard deviation) for students in technical and

---

[1] Empirical evidence suggests that intrinsic motivation (Fortier et al., 1996; Pintrich, 2003), social origins (Shavit and Blossfeld, 1993; Breen and Goldthorpe 1997; Bowles and Gintis 2002), parents' behaviour and expectations (Englund et al., 2004), teachers' expectations (Rosenthal and Jacobson, 1968; Rubie-Davies et al., 2006), and assessment practices (Machin and McNally, 2005) are pivotal in the learning process.

vocational schools, but no effect in academic schools. Our data show that, before the reform, about 40% of students in the areas considered had at least one 'education debt'. We argue that the size of the "marginal" group (i.e., students whose behaviour is affected by the reform) is large, and represents the large majority of the class if social externalities are at work.

We go beyond averages, and assess how the reform affected achievement across quantiles of the score distributions. We find that much of the variability in the effect is captured through the stratification by curricular track, with low within track differences across students. Most interestingly, we find more pronounced negative effects for females in non-academic schools. Results are robust to a number of sensitivity checks on the functional form and the estimation method adopted. Consistent with Betts and Grogger (2003), our first contribution is to show that higher standards coming with the threat of sanctions exacerbate inequality in the distribution of educational achievement, resulting in both winners and losers.

Second, we find that the reform induced adaptive behaviour of schools and, possibly, of their teachers, causing less stringent rules for grade progression (which in Italy is not based on standardised testing). We argue that this follows from the mandatory organisation of remedial summer courses, which impact importantly on school budgets. Despite the negative effects on scores, we find that the reform increased the fraction of students admitted to the next grade in non-academic schools. The same result holds up for academic schools, although significance in this case depends on the specification considered. We conclude that schools reacted to the reform by admitting to the next grade students who, with the old rules, would be given the 'educational debt'. We use this result to argue that the effects on scores mask important changes of schools' and teachers' behaviour. How the introduction of

remedial courses and exams affects outcomes is ambiguous, and depends on both schools'
and students' behaviour.

Finally, we perform back of the envelope calculations to infer the long term effects of
the reform on graduation rates. We exploit another reform that was the mirror image of the
intervention considered here. This reform was implemented in 1994/95, and introduced the
practice of social promotion by abolishing the same summer remedial exam considered
here. We use survey data from the Bank of Italy to set out the comparison of cohorts of
individuals aged 14 during the 1990s, this being the modal age for completing compulsory
schooling at the time. Our results point to no effect of the threat of grade retention
represented by the remedial exam on secondary school graduation rates. All in all, results
weigh against the effectiveness of the remediation policy considered here.

The remainder of the paper is organised as follows. A review of the literature is
presented in the next section. Institutional background and reform are discussed in Section
3. Section 4 presents data and evaluation design. Section 5 states the identifying conditions
required to ensure the causal validity of our findings. Results are reported in Section 6.
Conclusions and policy implication are discussed in Section 7. Additional materials are
presented in the Appendix.

2.   Literature review and implications for the empirical analysis

The desirability of grade retention policies to remediate poor performance is not
uncontroversial. The push for educational accountability in the recent years has brought this
problem back to the forefront. A large number of studies have looked into this issue, but
quasi-experimental evidence is still relatively scarce (notable exceptions are Jacob and
Lefgren, 2004 and 2009). Studies rigorously designed to control for selection bias fail to
demonstrate that grade retention is more beneficial than grade promotion, for both academic
and socio-emotional outcomes (see Jimerson, 2001). Our paper marks something of a

departure from this literature. In our setting the quasi-experimental comparison of students undergoing different progression rules reveals the increased threat of grade retention, not actual retention.[2]

As the reform we consider mandated remedial courses for low-achievers, the literature on the effects of remedial education is relevant to our study. Lavy and Schlosser (2005) consider a remedial intervention for high school students in Israel, finding a significant increase in matriculation rates. Calcagno and Long (2008) consider the impact of post-secondary remediation programmes in Florida, finding that mathematics and reading courses have mixed benefits on performance. Battistin and Meroni (2013) study the short term effects for low achieving students in Italian middle schools, and document positive results for math scores. De Paola and Scoppa (2014) consider remedial courses for college students in Italy, finding positive effects on credits.

The interplay between incentives faced by students and academic achievement is also important. Effort, total time devoted to study and engagement at school are important determinants of student learning (Bishop, 2004). Students choose which subject to focus on, and decide how much effort to put into each task. Depending on the incentives facing students, one may expect sizeable differences in decisions about effort. The available evidence suggests that determinants of effort vary a great deal across school tracks. Carbonaro (2005) finds that students in academic tracks exert substantially more effort than do students in non-academic tracks. Hastings *et al.* (2012) provide quasi-experimental evidence that school choice has sizeable effects on motivation and performance of low income and minority students. Raising the bar may have heterogeneous effects on effort, with negative consequences if standards move beyond a student's reach. Betts and Grogger

---

[2]To the best of our knowledge, the closest in spirit to our paper is the work by Belot and Vandenberghe (2014), who study the effects of the threat of grade retention introduced by a reform implemented for the French speaking community in Belgium finding no effects on achievement.

(2003), for example, find that high standards have significantly larger returns on test scores at the top end of the ability distribution. This result may be mediated by differential effects on effort, as students at the bottom end of the distribution may perceive themselves as losing ground and give up.[3]

Gender differences in the effort exerted may also be at work. Female students are more supportive of academic learning (see, for example, Carbonaro, 2005), implying that there may be positive externalities in schools with more girls. Lavy and Schlosser (2011) find that classes with more female students perform better, with higher scores for the most disadvantaged students. They argue that gender composition affects learning climate and inter-student relationships in the class. This should affect positively non-cognitive factors like motivation and concentration and thus, indirectly, learning. The effects of assessment mode on measured achievement also present significant gender differences. Machin and McNally (2005) document lower scores for girls after the introduction of continuous assessments in the UK National Curriculum. Gipps and Murphy (1994) show lower performance of girls in timed examinations because of higher anxiety. Powney (1996) reviews a number of studies documenting that the mode of assessment is a factor explaining the differential performance of male and female students.

These findings suggest that a sensible stratification for the empirical analysis is by gender and curricular track. This is what we will do in documenting our results.

3.   Institutional background and description of the reform

Secondary schools before the reform

Primary and secondary schooling in Italy is compulsory from ages 6 to 16, with three stages: 5 years of elementary school (*scuola elementare*), middle school covering grades 6-

---

[3] Betts and Grogger (2003) document differential effects of setting high standards by ethnicity, with lower returns on achievement for blacks. This is consistent with findings in Carbonaro (2005), who documents, *ceteris paribus*, lower effort from black students.

8 (*scuola media*), and secondary school (*scuola superiore*), which runs for 3-5 years. The reform we consider here affects students enrolled in latter stage of education, which is organized into academic (*liceo*) and technical and vocational tracks (*istituto tecnico* and *istituto professionale*, respectively). Teachers work in a highly regulated public sector, with virtually no risk of termination. Pay and promotion structure largely depend on seniority, and are independent of performance or conduct.

Data from multiple PISA waves consistently picture important geographic differences in academic performance at secondary school across areas (OECD, 2012). Students in the regions considered here (*Trentino*, *Veneto* and *Lombardia*) have fairly homogenous performance, and are among the best performing students in the OECD area. This is in sharp contrast with poor performance of Southern regions.

Until 2006/07 at the end of the school year (in June) students at secondary school could be retained, admitted to the next grade or admitted to the next grade with an 'educational debt' in one or more subjects. The 'debt' had to be cleared in the following years, with no clear deadline. Official figures show that about 42 percent of students in the country were given at least one 'debt', with just one out of four students recovering it by the end of the following year. Social promotion was the practice at work.

Table 1 presents descriptive statistics for year 2006/07 derived by the Ministry of Education, using areas considered for the main analysis. Separately for academic and technical and vocational schools, reported are the fraction of students retained, the fraction of students with at least one 'educational debt' at the end of the year, and a breakdown of debts assigned by subject. Figures here are for the second and third grades at secondary school, as this is the relevant age range for students in our sample.

Retention rates in non-academic schools are higher, and gender disparities are clear-cut. The difference across schools almost vanishes when it comes to 'debts', gender

remaining the most relevant factor. Over 40 percent of males are given at least one 'educational debt', and this figure is roughly the same across school types. The percentage of females with 'debts' is well below that of males, and females in non-academic schools are marginally worse than in academic schools. The breakdown by subject reveals that mathematics and foreign languages were, by far, the most problematic subjects. These numbers, which provide a representative picture for high schools in Northern Italy in the early 2000s, casted doubts on the learning effectiveness of the secondary system.[4]

Risk of grade retention and compliance with the reform

The reform implemented in 2007/08, named after the Minister of Education *Fioroni* in charge at the time, has *three* key features. First, under the new progression system - which is still in operation - students must clear all 'educational debts' before the beginning of the new year (mid-September). This is a requirement for admission to the next grade. Second, students with a 'debt' must attend remedial courses organized by the school during the summer, and take a remedial exam in early September (the assessment mode being decided by the school). In June students can be retained, admitted to the next grade, or must attend remedial classes over the summer in preparation for the exam in September. If this is not passed, grade retention is deliberated by the school council. Finally, schools are held responsible for the organization of summer courses and remedial classes during the year. Remedial classes for low performing students became mandatory and part of the school curriculum. This additional burden was not compensated by more financial resources transferred from the Ministry of Education to schools. It is well known that Italy marks a

---

[4] According to PISA 2000 data, Italy ranked above Spain, Portugal and Greece but far behind the 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 PISA 2003 results for 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 peers (OECD 2004). The overall picture was confirmed in the PISA 2006 survey (OECD 2007).

departure from the OECD average, as spending per student increased by only 0.5% in real terms between 1995 and 2010 (OECD, 2013).

Using its special autonomy from the central Italian government, the province of Trento in Northern Italy did not comply with the reform. Academic performance of students at PISA tests, as good as in top-performing OECD countries, and high appreciation for local secondary schools were at the origin of this decision.[5] In contrast with the rest of the country, schools in the province must organize courses for students with 'educational debts' at the *beginning of the new year* (thus leaving the summer free for teachers, students and their families), and lack of minimum academic standards in some subjects would not necessarily imply grade retention. Local principals, teachers, policy makers and politicians viewed the enforcement of summer courses as an excessive sanction, with possible negative externalities on students' learning and motivation in the secondary school cycle.

The reform introduced major changes to school accountability but *ceteris paribus* increased risk of grade retention for students outside the province of Trento. Progression of students with 'educational debts' before the reform, about 40% in the areas considered (see Table 1), is made conditional on additional instruction time – during the year and in the summer – and a final exam in September. Peers on the other side of the discontinuity border bear the cost of longer hours of remedial education, but with no obvious impact on grade progression. Making promotion conditional on the exam in September represents a real *threat*, not just additional time spent studying during the summer. Of course there may be additional (monetary and non-monetary) costs coming from the time spent studying during the summer, which extend to families and schools as well.

---

[5]For example, according to PISA 2006 the average score in mathematics is 462 in Italy, and 508 in the province of Trento.

Heterogeneous effects mediated by changes in effort and/or motivation are therefore one expected outcome. The reform may have represented a negative stimulus to commitment for students from lower social classes, for example because of their smaller opportunity cost of dropping out (Boudon 1974, and Breen and Goldthorpe, 1997). Differential effort across school subjects may be expected as well, with more investment where the threat is highest. Adaptive behaviour on the side of students is not the only possibility, of course, as schools and teachers may react to the administrative burden and increased costs by rationally exerting lower effort.

4.   Evaluation design and data

Selection of towns

The evaluation design considers towns close to the administrative border relevant for the reform. The province of Trento consists of few urban centres, and many rural suburbs and mountain villages with different population demographics. It is bordered by the regions of *Veneto* to the South-East, *Lombardy* to the West, and *Tyrol* to the North, as shown in Figure 1. The criteria followed to guide selection were: (i) presence of schools for each curricular track in secondary education, (ii) population size, and (iii) economic and occupation structure.

Condition (i) restricts the eligible set to the local capital, *Trento*, and few remaining urban centres. Excluding those with mountain villages and valleys in school catchments, which hardly are comparable to surrounding regions, the choice is reduced to three urban towns. The final selection includes the capital Trento and the towns of *Riva del Garda* and *Rovereto*, as shown in Figure 1. The area considered covers approximately two fifth of the province population.

For each of these three towns, we looked for the closest match among centres outside the administrative border (restricting the search to seventy kilometres from the boundary).

Here too the choice is constrained once conditions (ii) and (iii) are imposed. Given the small number of towns usable for the comparison, we didn't apply any statistical procedure to form pairs. We matched the capital *Trento* to *Bolzano*, the capital of South Tyrol (to the North). The town of *Riva del Garda* was matched to *Desenzano del Garda*, to the South. This is the only urban centre of comparable size, and the economies of both towns rely heavily on tourism. Finally, the relatively more industrial town of *Rovereto* was matched to the town of *Schio*, to the East.

The pairs constructed present very similar population size and age structure, as Figure 2 shows. We report, for each pair, time series of population size, share of working age population and labour force structure (defined as ratio of elderly to young workers) from census data and official publications from the Italian National Bureau of Statistics (ISTAT). Figure 3 and Figure 4 report demographics and selected labour market statistics. Educational attainment and occupational structure are remarkably similar across towns, for both young adults (aged below 30) and working-age population. Share of immigrants of secondary school age (13 to 18) is also comparable.

Selection of schools

We followed a *one-to-one* matching procedure, considering similar *public* schools located in each pair. Our choice is somewhat restricted by the population size of towns in the sample.[6] The selection was conducted by controlling for quantitative dimensions such as curricular track, school size as measured by trends in enrolment and school resources, as well qualitative dimensions as school `reputation'. The final sample includes 22 schools, almost 60% of existing secondary schools in the towns of interest. We selected 6 schools in

---

[6] For example, education received in academic schools (*liceo*) depends on the field of study (e.g., humanities or science). There is typically one academic institution per field in the towns considered, excluding private institutions and German institutions in *Bolzano*.

Trento and *Bolzano* (from a total of 9 and 7, respectively); 2 schools in *Riva del Garda* and *Desenzano del Garda* (from a total of 3 and 6); and 3 schools in *Rovereto* and *Schio* (from a total of 7 in both towns). Table 4 below shows that schools in the sample share the same outcomes with schools in the same towns not selected for this study (see the next section for details).

We focused on students in second and third grade at secondary school during the 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). We did so to ensure variability in the duration of enrolment at school across different regimes defined by the reform. The cohort dimension, however, proved statistically not important in the analysis, and will not be considered in what follows. Size of the working sample is presented in Table 2: number of students is 916 and 842 outside and inside the province of Trento, respectively, evenly distributed across curricular tracks.

Test scores and administrative data

Standardised testing for evaluation purposes is a recent development in Italy. The Italian *Istituto Nazionale per la Valutazione del Sistema dell'Istruzione* (INVALSI), a government agency charged with educational assessment, piloted voluntary assessments in 2008, which became mandatory for all schools from 2009. The lack of standardized scores for the period considered here forced us to collect our own data.

We combine school administrative databases (teachers' marks and information on promotion/retention) and information from two surveys designed for this study. The first is a small-scale survey shaped around PISA, and contains standardised scores for all students in our sample. The second survey – administered to parents – contains information on

12

household demographics such as education, labour market position, occupational standing, composition and learning resources at home. Data on test scores and demographics are referred to as "Main Sample". The test was developed from publicly released items from the first three PISA assessments available at the time that this research started (2000, 2003 and 2006). The procedure was supervised by experts of the Ministry of Education to guarantee comparability with PISA. The test was conducted at the beginning of 2008/09 (between October and November 2008).[7]

Standardized achievement is linked to teacher marks for past years at school, as well as to results at the final examination of middle school.[8] We also use administrative data from local government agencies and the Ministry of Education on retention rates for 2006/07, 2007/08, 2008/09 and 2009/10 in the areas considered. We construct longitudinal information for the *same* schools included in the main sample, and study changes in retention rates from before to after the reform over and above school- and grade-specific effects. The information for 2006/07, the only pre-reform year we were able to gather, is used to test for pre-existing differences between sampled schools in the areas considered (see Table 4 below).

5. Estimation

The evaluation design for the effects on scores sets up the cross-sectional comparison of outcomes for schools in the province of Trento, to outcomes in similar schools of adjacent areas. The identifying restriction is that schools near the administrative border would have had the same outcomes had the reform not been implemented (see, for example, Heckman

---

[7]The items were presented to students in three one-hour booklets, resulting in a three-hour session with 23 units for reading, 20 for mathematics and 19 for science. All students in our sample took the same tests.

[8] We investigated the possible sorting effects deriving from the choice of curricular track at high school across the areas considered in our analysis. We computed the average transition rates from middle school to *liceo* for the school year 2007/08 using official data from the Ministry of Education. This analysis pictured rather similar figures in the areas considered, with transition rates ranging between 29 and 34 percent.

and Vytlacil, 2007). We report estimates the following parametric regression from the "Main Sample":

$$y_i = \beta_0 + \beta_1 d_i + \beta_2 x_i + \varepsilon_i, \tag{1}$$

which is estimated separately by gender and curricular track. The variable y_i is score of student i (reading, math and science), d_i is an indicator for schools outside the province of Trento, and x_i are demographics reported in Table 3 below. Standard errors are clustered on class. Results are presented in Table 6.

We consider semi-parametric alternatives to this specification, which we use to check the sensitivity of our conclusions to the estimation method. First, the average effect of interest is estimated through a matching estimator. Matching is implemented using the propensity score obtained from a probit regression of "treatment" status on x_i. The two groups of students contrasted have similar distributions of observables, ruling out common support problems in the data. Second, we employ the propensity score to estimate the average effect through a weighting estimator (see Imbens, 2004). We additionally check the robustness of our findings by dropping few observations "off support" considering the propensity score metric.[9] The various sensitivity checks yield results equivalent to those obtained from the estimation of the parametric regression in (1), in terms of both point estimates and statistical significance. Because of this, we decided to report only parametric estimates while presenting the results in the following sections. Estimation results obtained using the weighting procedures are presented in Table A.1 of the Appendix.

In estimating quantile treatment effects from the "Main Sample", we fit the standard quantile regression counterpart of equation (1). Results from this analysis are presented in Figure 5. We check the sensitivity of our findings to alternative estimation methods

---

[9] Propensity score distributions across groups are reported in Figure A.4 of the Appendix. We dropped 56 observations off support.

employed in the programme evaluation literature (see, for example, Firpo, 2010), and results are very similar to those in the main text (see Figure A.1 of the Appendix).

When considering administrative data, the availability of longitudinal information from 2006/07 to 2009/10 allows us to consider a differences in differences strategy. For this reason, the variability employed in equation (2) below is different from that in equation (1). We estimate the following regression separately for academic and non-academic institutions:

$$y_{gjt} = \alpha_0(g,j,t) + \alpha_1 d_j p_t + \alpha_2 w_{gjt} + v_{gjt}, \tag{2}$$

where y_{gjt} is outcome (e.g., retention rate) in grade g for school j at time t, $\alpha_0(g,j,t)$ is shorthand for grade, school and time effects, p_t is dummy for post reform periods, d_j is an indicator for reform areas, and w_{gjt} is grade enrolment. Standard errors are clustered on school and enrolment cohort, and results are reported in Table 7.

6.   Results

6.1   Descriptive statistics

Students in the two areas share very similar demographics, as shown in Table 3. The following variables are considered: (i) *student demographics* (gender, age, dummy for foreign students, dummy for cohabitation with mother and father, proximity to school) (ii) *socio-economic background of the household* (father's age and education, mother's age and education, dummy for homemakers, dummy for mother or father unemployed), (iii) *household wealth and social-status* (occupational scale scores, material deprivation index).[10] Means and standard deviations are reported, stratifying observations by curricular track and policy status. All in all, the distribution of demographics is fairly balanced,

---

[10] The socio-economic status is measured using an Italian occupational stratification scale (De Lillo and Schizzerotto 1985) to proxy social standing of different jobs. The life style deprivation index (Whelan *et al.*, 2002) is an additive index based on the lack of 5 items in the household: television, car, DVD player, computer, internet access. Each individual item is weighted by the proportion of households possessing that item in Italy. Weights were derived from the SILC 2006 survey for Italy.

although some exceptions emerge (e.g., education of mothers). Results in Table 3, coupled with the evidence in Figure 2, Figure 3 and Figure 4, are our first attempt to corroborate the validity of the research design.

The "Main Sample" doesn't have information on scores *before* the reform. As such, it cannot be used to provide falsification tests in support of the identifying restriction underlying estimation of equation (1). We used PISA data to circumvent this problem. The Ministry of Education granted us access to information for 2003, 2006 and 2009 on the province where schools are located. However, the size of PISA sample ensures a small number of schools before (2003 and 2006) and after (2009) the reform, with different (academic and non-academic) institutions participating across survey waves. Moreover, as province represents a much broader geographic aggregation than the one considered here, additional confounding factors are likely to come into play.[11]

To overcome the lack of standardized scores before the reform, we use administrative data to show two additional results. First, promotion and retention rates in towns considered for this study are statistically the same for schools in and out of sample. This can be seen from columns (1) and (2) of Table 4, where reported is the coefficient γ_1 from regressions that consider *all schools* in towns selected for the years available:

$$y_{gjt} = \gamma_0(g,t) + \gamma_1 s_j + \gamma_2 w_{gjt} + \gamma_3 z_j + u_{gjt}. \tag{3}$$

---

[11] This said, and with the caveat of using an underpowered study, conclusions from this "PISA sample" are not against the hypothesis of parallel trends in scores pre-reform (results are not presented here for brevity; see Battistin and Schizzerotto, 2012). For descriptive purposes, we used this sample to investigate the distribution of key school inputs that may affect scores in the two areas. Student to teacher ratio and proportion of girls in the class are roughly comparable around the administrative border, and stable over time. In academic schools the teacher to pupil ratio is 8.65 and 8.01 inside and outside the province of Trento, respectively. The corresponding figures for non-academic schools are 6.49 and 7.27. The proportion of girls in sensibly lower in non-academic schools: about 44 percent in the province of Trento, and 41 percent outside this province. In academic schools, the proportion is about 66 percent and stable across areas. We conclude that the stratification by curricular track controls for different learning environments in the class (e.g., see Carbonaro, 2005, and Lavy and Schlosser, 2011).

The outcome y_{gjt} is regressed on a dummy for being a school in the sample, s_j, controlling for grade and time fixed effects, $\gamma_0(g,t)$, grade enrolment, w_{gjt}, and an indicator for academic institutions, z_j. Second, we keep only schools in the sample, and show that *before* the reform promotion and retention rates were similar either side of the administrative boundary. We estimate the following equation using data for 2006:

$$y_{gj2006} = \delta_0(g) + \delta_1 d_j + \delta_2 w_{gj2006} + \delta_3 z_j + \xi_{gj2006}, \tag{4}$$

where $\delta_0(g)$ is shorthand for grade fixed effects, and report the coefficient δ_1 in columns (3) and (4) of Table 4. We find weakly significant pre-reform differences for promotion rates, but no difference for retention rates.

Finally, Table 5 reports descriptive statistics for outcomes in equation (1). The mean difference between the two groups of areas is positive for academic schools, and negative for technical and vocational schools. Results are suggestive of disparities in achievement between treated and control schools, with negative differences for females from lower socio-economic backgrounds.

6.2 Effects on scores

Table 6 reports the average policy effects obtained by estimating equation (1). Results are presented by subject, gender and curricular track. Scores are standardized to have zero mean and unit variance, so that effects can be interpreted in points of standard deviation. Leaving aside significance, most estimates for academic schools have positive signs, while estimates for the remaining schools point to the opposite direction. Effects for academic schools are never different from zero. On the contrary we find large, negative effects for female students in technical and vocational schools, for both reading and science (-0.4 points). This sharp difference between curricular tracks is robust to the estimation method employed (see Table A.1 of the Appendix). Table 1 implies that number of students "at risk" was large, about 40% of the population in the areas considered. Combined with the

17

possibility of a social multiplier, most likely the reform affected the behaviour of the large majority of students in the class. This might explain the size of the effect documented, and the little within-class heterogeneity in the quantile effects presented below.[12]

To make up for lack of pre-reform achievement data, we estimated a value-added specification by conditioning on a proxy of student ability (in the spirit of Lavy *et al.*, 2012). For each student we considered teachers' marks at the end of the two semesters of the first and second grade at school, for both mathematics and language. We limited the analysis to these subjects as they are common across all curricular tracks. We also considered the final mark obtained at the end of the lower secondary school. This yielded a total of 8 to 12 marks per students, depending of the cohort of enrolment. Importantly, none of these marks is the result of standardized assessment. We stacked these indicators, together with scores collected through the main survey, and run gender-specific fixed effect regressions controlling for subject dummies (mathematics or science, vis-à-vis reading or language), nature of the indicator employed (test scores vis-à-vis administrative marks) and age when the indicator was measured. We used these regressions to predict each student's "fixed effect", which we included as an additional control in equation (1). Results in what follows are confirmed if we use only marks in periods pre-reform.

Results from this specification are reported in Table A.2 of the Appendix, and align well with the pattern in Table 6. In non-academic schools, only scores for female students

---

[12]We investigated the heterogeneity of results allowing the policy effect to vary across groups of students that were plausibly more at risk of grade retention under the new progression system. We exploited the variability across cohorts of enrolment in the original evaluation design (see Section 4), and selected only students enrolled for the first time in the school year 2006/07 (pre-reform). These students completed the first year at school under the old regime. We then run the regression in (1) only for this cohort of students, adding a dummy to identify students at risk of grade retention and its interaction with the area identifier. We experimented with alternative definitions of "students at risk". First, we considered students who were admitted to the second grade with at least one educational debt. Second, we focussed on students having a debt in mathematics, as we know from Table 1 that this was, by far, the case most frequently encountered situation. Finally, we labelled "at risk" those students who completed the lower secondary school with the minimum score. As expected, results from this set of regressions show that students more at risk of grade retention have actually lower test scores that their peers in the class. However, we rejected the hypothesis that results in Table 6 vary with risk.

are affected. The negative effects documented for reading and science are confirmed, although their magnitude is now attenuated (-0.3 points). Most results in Table 6 are confirmed for academic schools. The negative effect on reading scores for males, which was not statistically different from zero, is now more precisely estimated and significant at the conventional level.[13]

We tested whether the reform affected achievement across quantiles of the score distribution. For example, Figlio and Lucas (2004) show that high standards in the class have the largest effects on achievement for students mismatched with the average ability of their peers. Figure 5 reports quantile treatment effects (QTEs) for the various groups considered, along with the corresponding 95 per cent confidence intervals. For academic schools all figures are not statistically different from zero, pointing to homogeneous effects. The pattern for non-academic schools also supports the hypothesis of constant effects, but with a negative shift in reading and science scores for female students. Overall, the evidence documented suggests much lower within track variability in policy effects than between tracks. A possible explanation for this result in that school tracking creates homogeneous classes with respect to ability and family background.[14]

---

[13] Conclusions on scores are confirmed using the "PISA sample" (with the caveat about precision acknowledged above). Moreover using PISA data we find no significant effects on household spending for education, the bulk of which - given the public school system in Italy - consists in fees paid for private tutoring. According to our findings, households did not react to the reform providing extra support to students facing an increased threat of grade retention. In contrast, we find that the amount of extra time spent by students learning subjects outside of normal school hours increased after the reform. Of course this effect could be explained as mechanical consequence of the intervention itself, being the provision of remedial classes for low achieving students compulsory on the school side. However, our results show that much of the action took place in non-academic schools, while in academic schools the provision of remedial classes is unaffected by the reform. This evidence supports the idea that the reform lowered the "safety" of the most struggling students, imposing substantial extra work loads only for those from low socio-economic backgrounds.

[14] As for average effects, we checked the sensitivity of QTEs to the specification and the estimation method adopted. Figure A.1 of the Appendix is the analogue of Figure 5, but is obtained using the semi-parametric procedure suggested by Firpo (2010). We additionally derive the analogue of Figure 5 when quantile regressions include student fixed effects (see Figure A.2 of the Appendix).

6.3 Effects on retention rates

Moving from the evidence discussed in the previous section, we use administrative records to study the effects on retention rates. Results in Table 7 are derived from micro data by school and grade. We report separate results for academic and non-academic schools. In the latter case, after a matriculation exam at the end of the third year students can attain a formal qualification that enables to practice an occupation. Because of this, we consider data across curricular types for grades without the matriculation exam (grades 1 to 4, excluding the third grade in vocational schools).

We first consider in columns (1) and (4) students for whom the final status (retention or promotion) is determined in June. All remaining students have an 'educational debt' (schools in the province of Trento), or are mandated to summer courses and the remedial exam in September (schools outside the province of Trento). Results for promotion rates, as determined in June, are reported in columns (2) and (5). Columns (3) and (6) report retention rates in June. Equation (2) is first estimated using schools in the "Main Sample" (Panel A), and then including all schools in the towns selected (Panel B).

We find a significant increase in students whose status is determined in June which is driven by higher promotion rates – see columns (2) and (5).[15] These results imply that students admitted to the next grade in the post reform period include those who, before the reform, would have the educational debt. Effects for retention rates are never significant across all columns – see columns (3) and (6). Consistent with what documented in the first two columns of Table 4, estimation results for all schools in towns yield identical conclusions.

---

[15] Results for non-academic schools are robust to clustering on school only. Results for academic schools in this case become marginally significant.

Results in Table 7 are at odds with the evidence documented for scores in the "Main Sample". Despite the negative effects on scores in non-academic schools, these picture an increase in number of students admitted to the next grade. The same comment extends to academic schools, where no effects on test scores are detected. These findings are consistent with changes in schools' behaviour in the new regime. The reform introduced additional administrative burdens related to the organisation of remedial courses and the exam in September, leaving school budgets substantially unaffected. Schools reacted by admitting to the next grade those students who would otherwise have an 'educational debt'.

6.4   Long-term effects

We use the variability introduced by a *different* reform that affected upper secondary schools during the 1990s, and has many similarities with the intervention considered here. Remedial exams for low performing students were introduced in Italian schools in 1923. In 1994/95 the exam was abolished, and replaced by *debito formativo*. The last cohort of high school graduates before the 1994/95 intervention consists of students born in 1976. For these students, remedial exams had substantially the same format as the exams introduced with the 2007/08 reform. In this sense, the former reform represents the mirror image of the latter. Contrary to the most recent reform, however, the former intervention did not make remedial courses mandatory in the formative plan, and therefore did not constrain school financial resources.

We compare cohorts of students born before and after 1976, and use the available longitudinal dimension to look at their outcomes. Figure 6 reports official statistics for retention rates by grade and curricular type published by ISTAT. The reform year, marked with a dot in each profile, is associated with negative spikes at all grades, especially for vocational schools. However, leaving aside adjustments around the reform year, the time series does not present any evident structural break from the change to the new progression

rule. As a result, the risk of grade retention does not seem to have an impact on retention rates. Interestingly, we can't exclude that this finding was also in the information set of students, schools and teachers in the cohorts considered for our analysis.

Figure 7 maintains the same design, but considers graduation rates at the end upper secondary school (at age 18/19). Due to the lack of administrative information, graduation rates are computed using the 2010 wave of the Survey of Household Income and Wealth from the Bank of Italy (Labour Force Survey data yield similar results). No clear discontinuity emerges from the figure, for both male and female students.[16]

The 1994/95 and the 2007/08 reforms can be considered as exogenous shocks to the risk of grade retention; in addition, the latter reform impacted negatively on school budgets by making the organisation of remedial courses mandatory. The absence of clear effects of the 1994/95 reform, as opposed to the effects documented in Table 6, represents corroborative evidence in favour of the role played by changes in school inputs in the most recent reform.

7.   Discussion and conclusions

Heterogeneous effects of the reform on student effort could be one driver of our results. Children of advantaged social classes show a more positive attitude towards learning and feel more confident in their performance (Boudon, 1974). Lower class families have strong risk aversion towards investment in education because of low expected performance (Breen and Goldthorpe, 1997). It follows that students from these families, which are over-

---

[16] In principle, by adopting the same approach we could shed light of longer term effects on university completion and wages. However, cohorts born in the late 70s were also the most affected by a subsequent reform of the university system (see, for instance, Cappellari and Lucifora, 2009), that resulted for those cohorts into higher graduation rates and, therefore, indirect effects on labour market outcomes (see Figure A.3 of the on-line Appendix). The discontinuity in Figure 7 for cohorts born in the early 80s follows from another reform that, in the late 1990s, introduced new rules for the matriculation exam at secondary school (*Legge 10 November 1997, n. 425*).

represented in non-academic schools, may be reluctant to additional commitment when payoff is uncertain.

The negative effect on scores is stronger for female students, and more pronounced in reading and science. Female students in non-academic schools perform, on average, well below their male peers. This applies to mathematics in particular, which may be perceived as the subject with the highest return and for which acceptable performance raises the chances of grade progression. As girls are more risk-averse to school failures (Borghans *et al.*, 2009), the reform may have induced them to invest more in mathematics and exert less effort in other subjects. The result is exacerbated by secondary school tracking, which creates homogeneous classes according to ability and social background and reduces peer effects (see Zimmer, 2003, Hanushek and Wößmann, 2006). Students perform better if their peers are high achievers, and good peers can act as a buffer against deviant behaviour. Social stigmas may therefore be less important in schools where a non-negligible share of the population acts lowering effort. For example, Figlio and Lucas (2004) document sharp effects of grading standards for low-ability students when classroom ability is relatively high, supporting the idea that high standards improve the achievement of students mismatched with the typical ability level of their peers.

We have discussed how the reform imposed additional administrative burdens, with no adjustment to financial resources transferred to schools (OECD, 2013). We have provided evidence that schools reacted by admitting to the next grade students whom, in the pre-reform period, would have an 'educational debt'. If schools lowered their standards to avoid the organization of summer courses and the remedial exam in September, the negative effect on test scores can be driven also by adaptive behaviour of students. A competing explanation comes from adaptive behaviour of teachers. If principals and the governing bodies of schools decide to lower the standards required for progression, teachers could rationally

exert lower effort and teach at that standard, with an impact on academic achievement. 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 (Akerlof and Cranton, 2003). Adaptive behaviour of both students and teachers may induce a vicious circle in non-academic tracks in particular, as students may adjust their commitment to academic standards because of the low opportunity cost of dropping out (Breen and Goldthorpe, 1997). As teachers work in a highly regulated public sector without pay incentives and virtually no risk of sanctions, the temptation to adjust effort to the learning attitude of the class may be high.

These findings raise a number of questions, including what can be done to reinforce positive attitudes towards learning for the most disadvantaged students, and whether more informed decisions about retention or promotion should rely on standardized testing. We hope to address these questions in future work.

References

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

Battistin, E. and Schizzerotto, A. (2012). Threat of Grade Retention, Remedial Education and Student Achievement: Evidence from Upper Secondary Schools in Italy. IZA Discussion paper 7086.

Battistin, E. and Meroni, E.C. (2013). Should We Increase Instruction Time in Low Achieving Schools? Evidence from Southern Italy. IZA Discussion paper 7437.

Belot, M. and Vandenberghe, V. (2014). Evaluating the "Threat" Effects of Grade Repetition. Exploiting the 2001 Reform by the French-Speaking Community of Belgium. *Education Economics*, Volume 22, No. 1, pp. 73-89.

Betts, J.R. and Grogger, J. (2003). The impact of grading standards on student achievement, educational attainment, and entry-level earnings. *Economics of Education Review*, 22(4), pp. 343-352.

Bishop, J.H. (2004). Drinking from the fountain of knowledge: Student incentive to study and learn-externalities, information problems and peer pressure. *CAHRS Working Paper 04-15*. Ithaca, NY: Cornell University, School of Industrial and Labor Relations, Center for Advanced Human Resource Studies.

Borghans, L., Golsteyn, B. H., Heckman, J.J. and Meijers, H. (2009). Gender differences in risk aversion and ambiguity aversion, *Journal of European Economic Association*, 7, 2-3, pp. 649-658.

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

Bowles, S., and Gintis, H. (2002). The Inheritance of Inequality, *Journal of Economic Perspectives*, 16, 3, pp. 3-30.

Breen, R. and Goldthorpe, J. (1997). Explaining educational differentials: Towards a formal rational action theory. *Rationality and Society,* 9(3)**,**pp. 275-305.

Calcagno, J.C. and Long, B.T. (2008). The Impact of Postsecondary Remediation Using a Regression Discontinuity Approach: Addressing Endogenous Sorting and Noncompliance. *NBER Working Paper No. 14194*.

Cappellari, L. and Lucifora, C. (2009). The "Bologna Process" and college enrolment decisions, *Labour Economics*, 16, 6, pp. 638-647.

Carbonaro, W. (2005). Tracking, Students' Effort, and Academic Achievement. *Sociology of Education*, 78(1), pp. 27–49.

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

De Paola, M., and Scoppa, V. (2014). The effectiveness of remedial courses in Italy: a fuzzy regression discontinuity design, Journal of Population Economics, 27(3), pp. 65–386.

Englund, H.M., Luckner, A.E., Whaley, G.J.L. and Egeland, B. (2004). Children's Achievements in Early Elementary School: Longitudinal Effects of Parental Involvement, Expectations, and Quality of Assistance, *Journal of Educational Psychology*, 96, 4, pp. 723-730.

Figlio, D.N. and Lucas, M.E. (2004). Do High Grading Standards Affect Student Performance? *Journal of Public Economics*, 88, pp. 1815-1834.

Firpo, S. (2010). Efficient Semiparametric Estimation of Quantile Treatment Effects. *Econometrica*, 75(1), pp. 259-276.

Fortier, M.S., Vallerand, R.J. and Guay, F.(1996). Academic motivation and School Performance: Toward a Structural Model, *Contemporary Educational Psychology*, 20, 3,pp. 257-274.

Gipps, C. and Murphy, P. (1994). A Fair Test? Assessment, Achievement and Equity. Open University Press.

Hanushek, E.A. and Wössmann, L. (2006). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *Economic Journal*, 116(510), pp. 63-76.

Hastings, J.S., Neilson, C.A. and Zimmerman, S.D. (2012). The Effect of School Choice on Intrinsic Motivation and Academic Outcomes. NBER Working Paper No. 18324.

Heckman, J.J. and Vytlacil, J.E. (2007). Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation. In *Handbook of Econometrics*, Chapter 70, Volume 6, Part B, pp. 4779-4874.

Imbens, G.W. (2004). Nonparametric Estimation of Average Treatment Effects under Exogeneity. a Review. *Review of Economics and Statistics*, 86(1), pp. 4-29.

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

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

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

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

Lavy, V. and Schlosser, A. (2011). Mechanisms and Impacts of Gender Peer Effects at School. *American Economic Journal: Applied Economics*, 3(2), pp. 1-33.

Lavy, V., Silva, O. and Weinhardt, F. (2012). The Good, the Bad and the Average: Evidence on the Scale and Nature of Ability Peer Effects in Schools. *Journal of Labor Economics*, Volume 30, No. 2, pp. 367-414.

Machin, S. and McNally, S. (2005). Gender and Student Achievement in English Schools. *Oxford Review of Economic Policy*, 21(3), pp. 357-372.

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

OECD (2012), *Programme for international student assessment (PISA)*. Results from Pisa 2012 – Country Note, Italy.

OECD (2013), *Education at a Glance*, Country Report-Italy, Paris.

Pintrich, P.R. (2003). A Motivational Science Perspective on the Role of Student Motivation in Learning and Teaching Context, *Journal of Educational Psychology*, 95, 4, pp. 667-686.

Powney, J. (1996). Gender and Attainment - A Review. Edinburgh: Scottish Council for Research in Education.

Rosenthal, R. and Jacobson, L. (1968). *Pygmalion in the Classroom*, New York, Holt, Reinhart & Winston.

Rubie-Davies, C., Hottie, J. and Hamilton, R. (2006). Expecting the best for Students: Teacher Expectation and Academic Outcomes, *British Journal of Educational Psychology*, 76, 3, pp. 429-444.

Shavit, Y., and Blossfeld, H.P. (eds) (1993). *Persistent Inequality. Changing Educational Attainments in Thirteen Countries*, Boulder (CO), Westview Press.

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

Whelan, C., Layte, R. and Maitre, B. (2002). Multiple deprivation and persistent poverty in the European Union. *Journal of European Social Policy,* 12(2)**,**pp. 91-105.

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

Table 1. Descriptive statistics on student performance for the school year 2006/07 (pre-reform).

	Grade 2		Grade 3	
	Males	**Females**	**Males**	**Females**
Academic Schools				
% of students retained	8.9	6.4	6.8	5.4
% of students with debts	42.5	27.2	47.5	28.8
in Italian language	11.1	12.4	8.3	8.5
in Mathematics	62.0	58.1	43.5	47.2
in Foreign Languages	32.9	41.6	24.0	41.2
Technical and Vocational Schools				
% of students retained	16.2	14.2	14.5	11.3
% of students with debts	40.5	32.7	42.6	35.4
in Italian language	15.4	11.1	17.3	10.1
in Mathematics	46.8	45.9	51.9	47.6
in Foreign Languages	36.2	33.1	39.1	37.7

Notes. Administrative data from the Ministry of Education for schools located in the provinces considered (*Trento*, *Bolzano*, *Verona* and *Vicenza*). Figures are presented by gender, curricular type (academic and technical or vocational) and grade. See Section 3 for details.

Table 2. Sample size: students, classes and schools (Main Sample).

	Males		Females	
	on	**off**	**on**	**off**
Academic Schools				
Students	168	146	281	205
Classes	23	20	23	20
Schools	5	4	5	4
Technical and Vocational Schools				
Students	266	269	201	222
Classes	28	28	28	28
Schools	6	7	6	7

Notes. Sample size by gender, curricular type (academic and technical or vocational) and policy status of the area where the school is located (on and off stand for outside and inside the province of Trento, respectively). See Section 4 for details.

Table 3. Descriptive statistics and balancing tests for the covariates considered (Main Sample).

VARIABLES	(1) Mean	(2) SD	(3) Mean	(4) SD	(5) Diff	(6) Mean	(7) SD	(8) Mean	(9) SD	(10) Diff
	Academic Schools					**Technical and Vocational Schools**				
	Policy On		**Policy Off**			**Policy On**		**Policy Off**		
Student Demographics										
Age	15.837	0.707	15.755	0.671	0.625	16.005	0.807	15.963	0.788	0.774
Male	0.374	0.484	0.416	0.494	0.511	0.570	0.496	0.548	0.498	0.832
Born abroad	0.547	0.498	0.443	0.498	0.495	0.124	0.330	0.122	0.328	0.968
Cohabitating with mother and father	0.130	0.337	0.081	0.274	0.082	0.192	0.394	0.177	0.382	0.642
School in the place of residence	0.517	0.500	0.456	0.499	0.331	0.261	0.440	0.265	0.442	0.952
Household Demographics										
Father's age	48.625	5.448	49.061	4.727	0.235	47.963	5.879	47.896	5.639	0.883
Mother's age	45.404	4.586	46.145	4.371	0.043	44.659	4.999	44.562	5.132	0.798
Father's education: Secondary	0.530	0.500	0.530	0.500	0.996	0.545	0.499	0.532	0.500	0.730
Father's education: Tertiary	0.280	0.450	0.332	0.472	0.236	0.083	0.276	0.097	0.296	0.443
Mother's education: Secondary	0.581	0.494	0.540	0.499	0.281	0.528	0.500	0.632	0.483	0.003
Mother's education: Tertiary	0.232	0.423	0.341	0.475	0.012	0.093	0.291	0.077	0.268	0.414
Mother is housewife	0.160	0.367	0.171	0.377	0.642	0.226	0.419	0.229	0.421	0.928
Mother or father unemployed	0.018	0.135	0.012	0.109	0.495	0.030	0.196	0.026	0.160	0.762
Wealth and Social Status										
Life-stype deprivation index	0.300	1.083	0.259	1.006	0.521	0.535	1.433	0.583	1.478	0.722
Occupational prestige scale	51.391	19.314	56.750	19.393	0.011	40.221	17.803	39.524	17.159	0.602
Number of students	449		351			467		491		

Notes. Reported are descriptive statistics for the covariates used as controls in the analysis, by curricular type (academic and technical or vocational) and policy status. Columns (5) and (10) report p-values for the statistical significance of the difference between columns (1) and (3) and (6) and (8), respectively. Standard errors are clustered on class (see Section 6.1 for details). *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 4. Balancing tests (administrative data).

	All Schools		Schools in Sample	
	(1)	(2)	(3)	(4)
	Admitted	**Retained**	**Admitted**	**Retained**
In Sample	0.0024	-0.0022		
	(0.014)	(0.007)		
Policy on			-0.0387*	0.0119
			(0.023)	(0.011)
Baseline			0.5033	0.1303
Observations	516	516	85	85

Notes. Estimates of equation (3) are reported in columns (1) and (2) for promotion and retention rates, respectively, using all years from 2006/7 to 2009/10. All schools in the towns selected are considered, and observations correspond to school/grade/year outcomes. Estimates of equation (4) are reported in columns (3) and (4) for promotion and retention rates, respectively, using the only pre-reform year available, 2006/07. Only schools in the sample are considered, and observations correspond to school/grade/year outcomes. Baseline figures represent the average of outcomes in 2006/07, across grades at secondary school, in towns not affected by the reform. Standard errors are clustered on school. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 5. Descriptive statistics for scores (Main Sample).

| | | **Males** | | **Females** | |
		On	**Off**	**On**	**Off**
Academic Schools					
Reading	Mean	500.670	509.438	521.042	512.092
	SD	91.090	102.069	98.128	101.102
Math	Mean	540.312	523.952	498.814	492.304
	SD	79.690	91.804	91.999	89.734
Science	Mean	545.119	523.044	520.717	513.828
	SD	86.278	83.872	86.833	85.055
Technical and Vocational Schools					
Reading	Mean	411.271	418.770	407.558	447.144
	SD	80.488	85.151	85.194	93.376
Math	Mean	480.705	484.796	413.787	417.886
	SD	113.330	96.482	115.081	117.043
Science	Mean	453.240	460.934	418.407	448.835
	SD	104.739	101.633	89.128	86.668

Notes. Summary statistics by gender, curricular type (academic and technical or vocational) and policy status of the area where school is located.

Table 6. Average effects on scores (Main Sample).

| | **Reading** | | **Mathematics** | | **Science** | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
	Males	**Females**	**Males**	**Females**	**Males**	**Females**
Academic Schools						
	-0.1931	0.1751	0.0657	0.0684	0.1974	0.1287
	(0.125)	(0.138)	(0.140)	(0.127)	(0.140)	(0.107)
Technical and Vocational Schools						
	0.0286	-0.4320***	-0.0757	-0.0119	-0.0041	-0.4330***
	(0.132)	(0.126)	(0.141)	(0.110)	(0.126)	(0.131)

Notes. Results from equation (1) for separate regressions by gender and curricular type (academic and technical or vocational). All regressions include the variables in Table 3, and outcomes are standardized to have zero mean and unit variance. Robust standard errors, clustered on class, are reported in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 7. Effects for retention and promotion rates (administrative data).

	Academic Schools			Technical and Vocational Schools		
	(1)	(2)	(3)	(4)	(5)	(6)
	In June	**Admitted**	**Retained**	**In June**	**Admitted**	**Retained**
	Panel A. Schools in the sample					
Effect	0.1409***	0.1359***	0.0049	0.0716***	0.0666***	0.0050
	(0.040)	(0.041)	(0.007)	(0.022)	(0.024)	(0.014)
Observations	144	144	144	196	196	196
	Panel B. All schools					
Effect	0.1245***	0.1201***	0.0044	0.0898***	0.0795***	0.0103
	(0.040)	(0.041)	(0.007)	(0.015)	(0.017)	(0.012)
Observations	176	176	176	340	340	340

Notes. Estimates of equation (2), by curricular type, using administrative data (Italian Ministry of Education and local agencies). Promotion and retention rates for time/grade/schools cells for years 2006/07 (pre-reform), 2007/08, 2008/09 and 2019/10 (post-reform). Columns (1) and (4) refer to students whose final status (retention or promotion to the next grade) is determined in June. Columns (2) and (5) refer to students admitted to the next grade in June. Columns (3) and (6) report retention rates in June. Estimation results in Panel A are limited to schools in the sample. Panel B considers all schools in towns selected. Standard errors, in parentheses, are clustered on school and enrolment cohort. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Figure 1. Evaluation design.
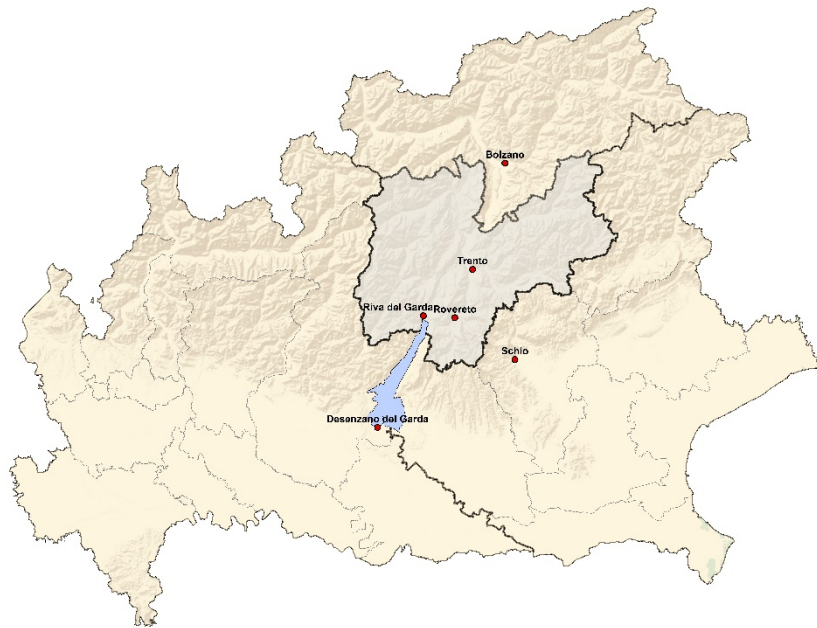
Notes. The province of *Trento* (Northern Italy) is the only province not affected by the reform. Four provinces are considered: *Trento*, *South Tyrol* (North of Trento), *Verona* (South-West) and *Vicenza* (South East). An Italian province is a territory administratively similar to a US county. We matched the capital *Trento* to the capital of Tyrol, *Bolzano*. The town of *Riva del Garda* was matched to *Desenzano del Garda*, in the province of Verona. The town of *Rovereto* was matched to *Schio*, in the province of Vicenza. See Section 4 for details.

Figure 2. Population size and age structure in the towns selected.
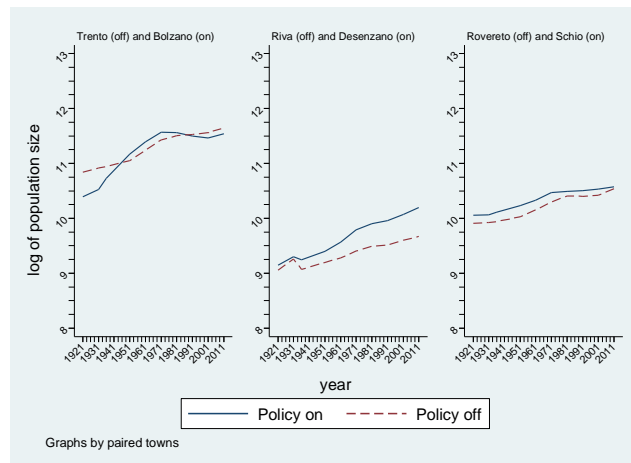
Panel A. Population size
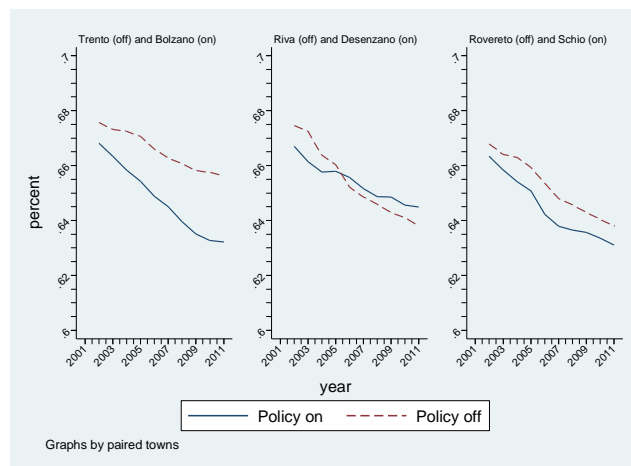
Panel B. Working age population, 15-64
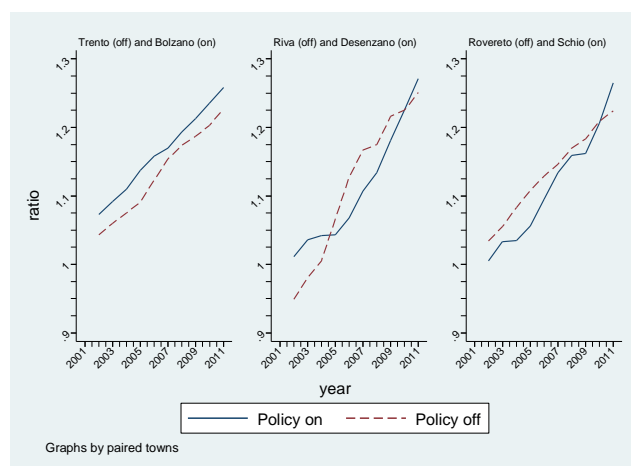
Panel C. Population 40-64 over population 15-39

Notes. Panel A is obtained from census data (1921 to 2011). Panel B and Panel C use data from the Italian National Bureau of Statistics (ISTAT) from 2002 to 2011. The three pairs of towns in Figure 1 are considered (*Trento* vs *Bolzano*, *Riva del Garda* vs *Desenzano del Garda*, and *Rovereto* vs *Schio*). See Section 4 for details.

Figure 3. Educational attainment in the towns selected.
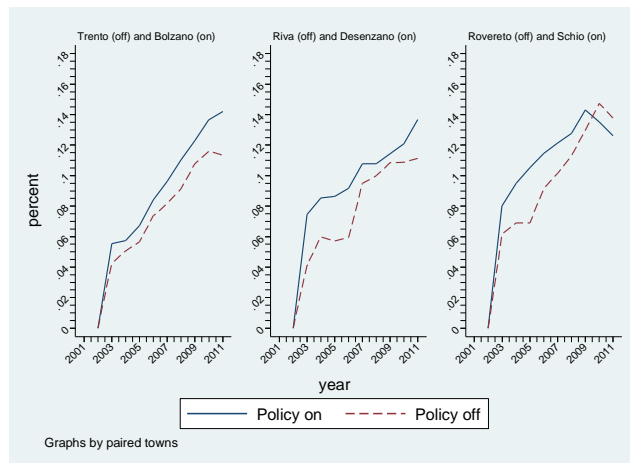
Panel A. Percent of immigrants aged 13 to 18
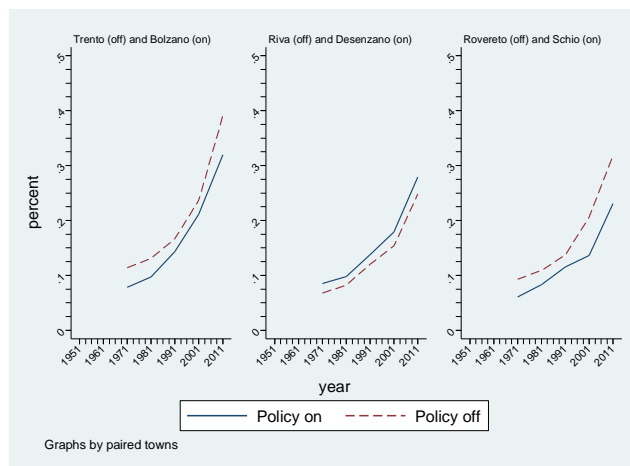
Panel B. Percent of young adults with tertiary education
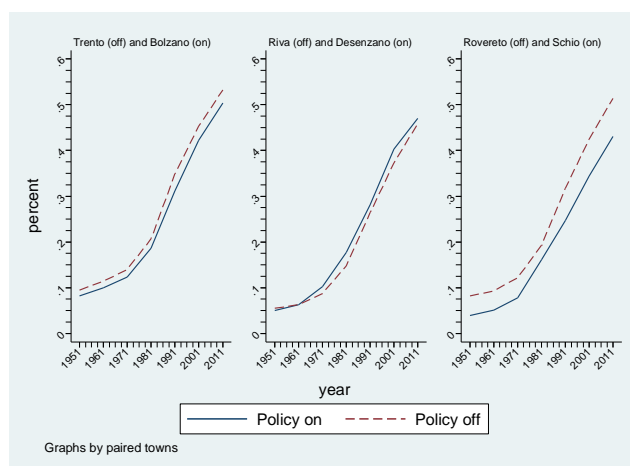
Panel C. Percent with at least high school diploma

Notes. Panel A uses data from the Italian National Bureau of Statistics (ISTAT) from 2002 to 2011. Panel B and Panel C use census data (1951 to 2011). The three pairs of towns in Figure 1 are considered (*Trento* vs *Bolzano*, *Riva del Garda* vs *Desenzano del Garda*, and *Rovereto* vs *Schio*). See Section 4 for details.

Figure 4. Labor market characteristics in the towns selected.
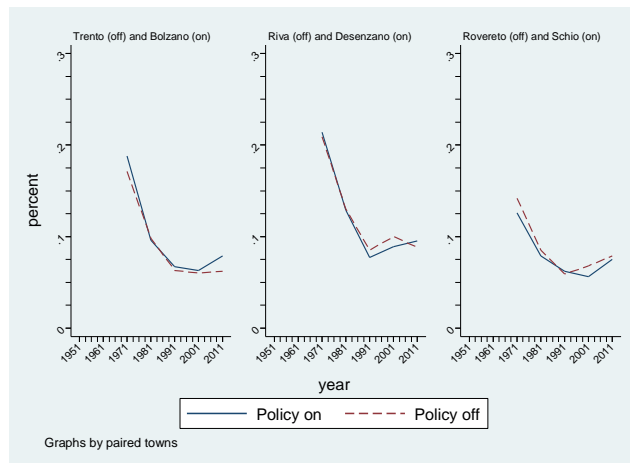
**Panel A**. NEET population aged 15 to 29
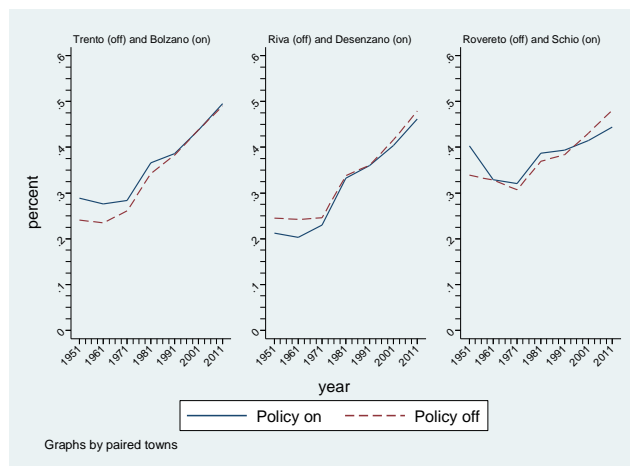
Panel B. Female participation rate
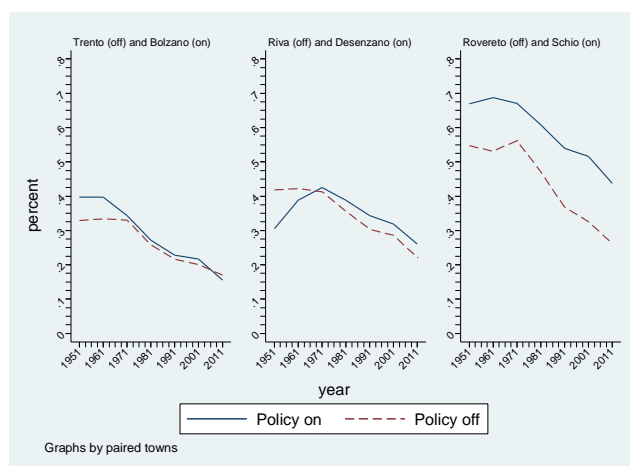
Panel C. Economic activity: percent working in secondary sector

**Notes**. Panel A, Panel B and Panel C use census data (1951 to 2011). The three pairs of towns in Figure 1 are considered (*Trento* vs *Bolzano*, *Riva del Garda* vs *Desenzano del Garda*, and *Rovereto* vs *Schio*). See Section 4 for details.

Figure 5. Quantile treatment effects for scores (Main Sample).
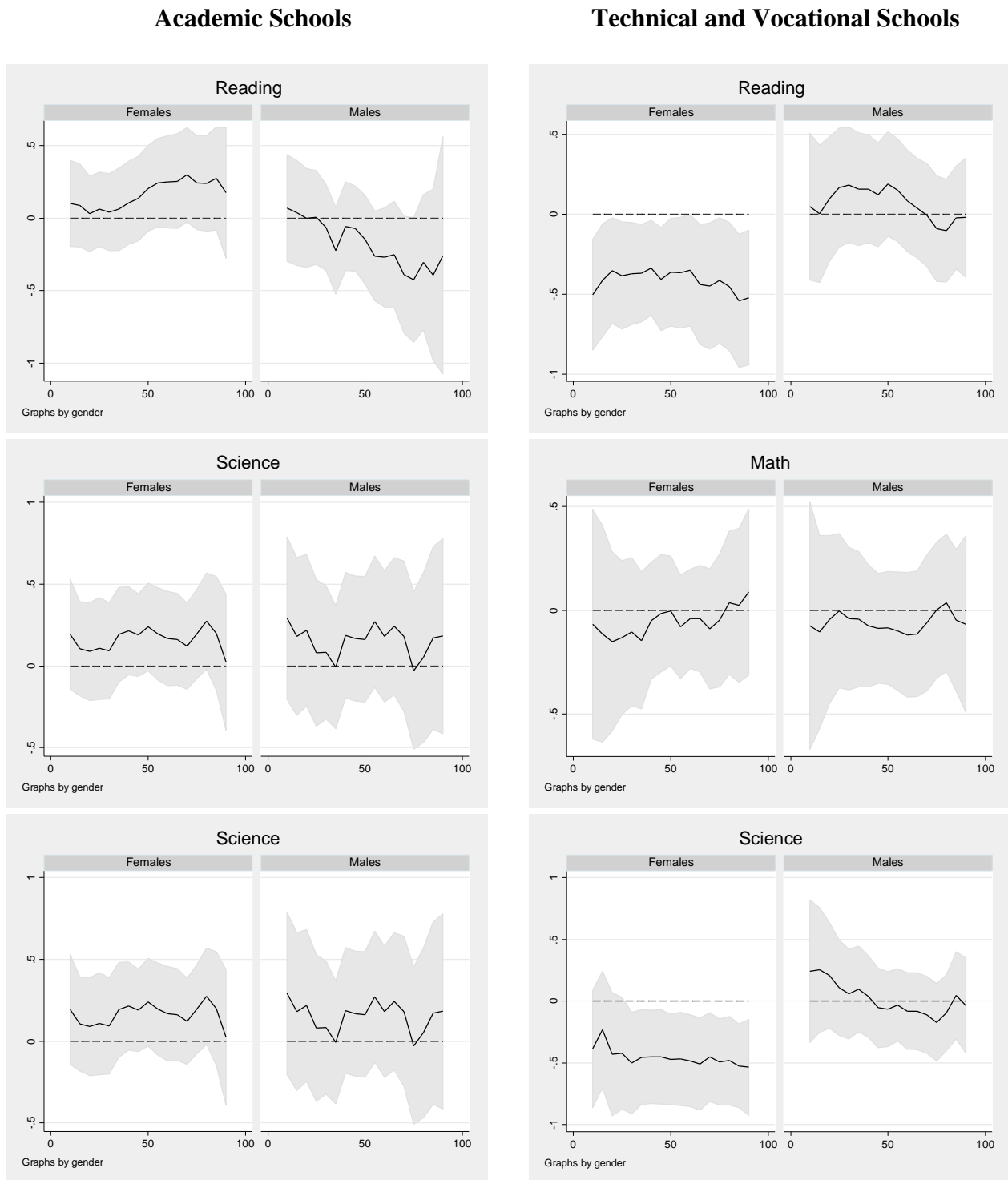
<div align="center">Academic Schools   Technical and Vocational Schools</div>


Notes. Reported are estimates of quantile regressions for reading, mathematical and scientific literacy. Results were obtained from separate regressions by gender and curricular type (academic and technical or vocational). All regressions control for variables in Table 3, and outcomes are standardized to have zero mean and unit variance. Bootstrap standard errors based on 200 replications and clustered on class were used to compute 95% confidence intervals.

Figure 6. Retention rates by cohort of birth (1994 reform).

Graphs by School type
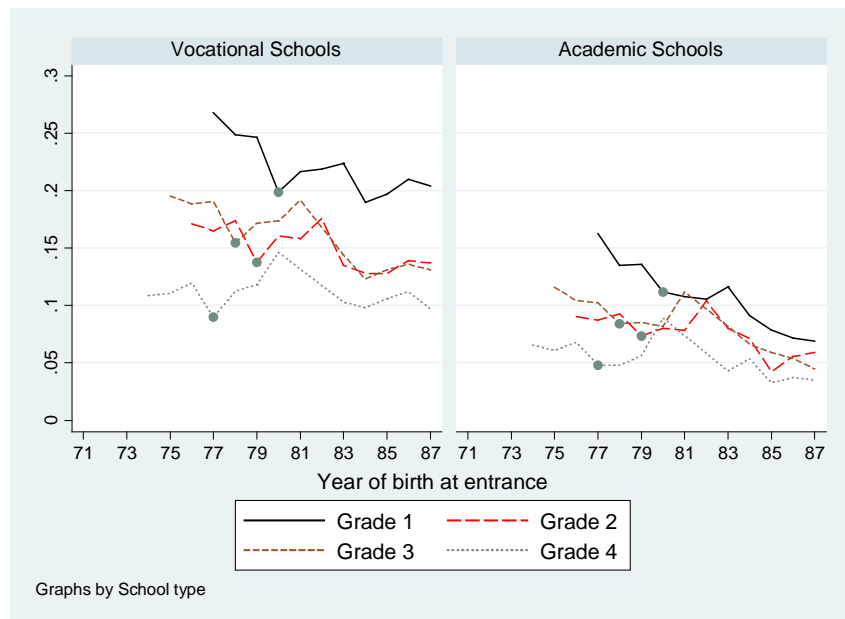
Notes. Reported are retention rates by cohort of enrollment at secondary school (e.g. 78 stands for "born in 1978" and enrolled at age 14), by school grade. Only regions in the Northern/Central part of Italy are considered. The reform year is marked with a grey dot. Source: ISTAT (Italian National Bureau of Statistics, before the school year 1997/98) and MIUR (Ministry of Education, from the school year 1998/99). See Section 6.4 for details.

Figure 7. Percentage of secondary school graduates by cohort of birth (1994 reform).

Graphs by gender
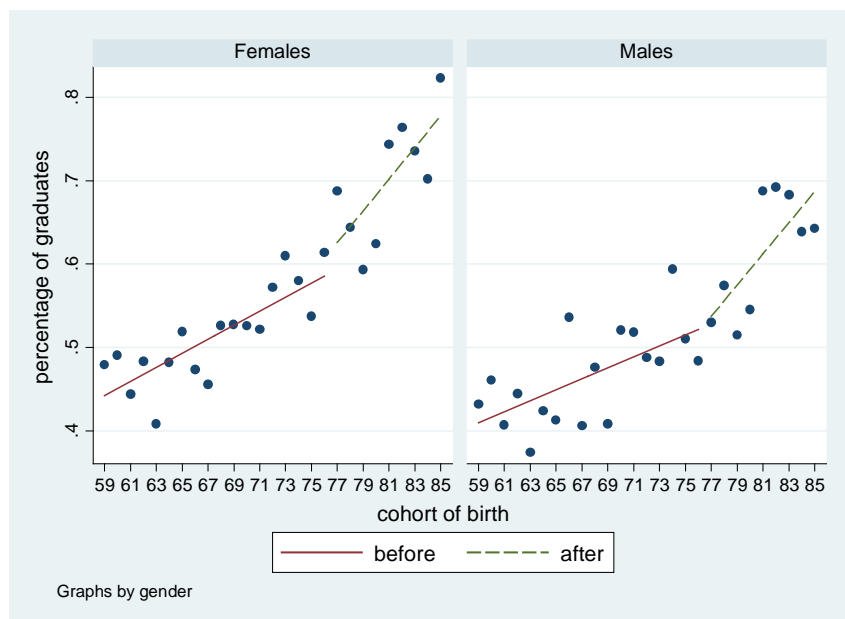
Notes. Reported is the percentage of secondary school graduates by cohort of birth, computed from the 2010 Bank of Italy Household Survey on Income and Wealth. Cohorts born before 1976 are not affected by the 1994/95 reform of secondary school system (see Section 6.4 for details).

Appendix

(not for publication)

Table A.1. Average effects on scores (sensitivity analysis - Main Sample).

Academic Schools

	Reading		Mathematics		Science	
	Males	Females	Males	Females	Males	Females
Effect	-0.2143	0.1941	0.0325	0.0798	0.1698	0.1296
	(0.1613)	(0.1315)	(0.1610)	(0.1316)	(0.1924)	(0.1018)
Baseline	521.3540	505.7525	532.5818	491.2104	528.2151	512.9147
	(12.6388)	(7.8504)	(11.3752)	(6.9062)	(11.7942)	(7.1733)

Technical and Vocational Schools

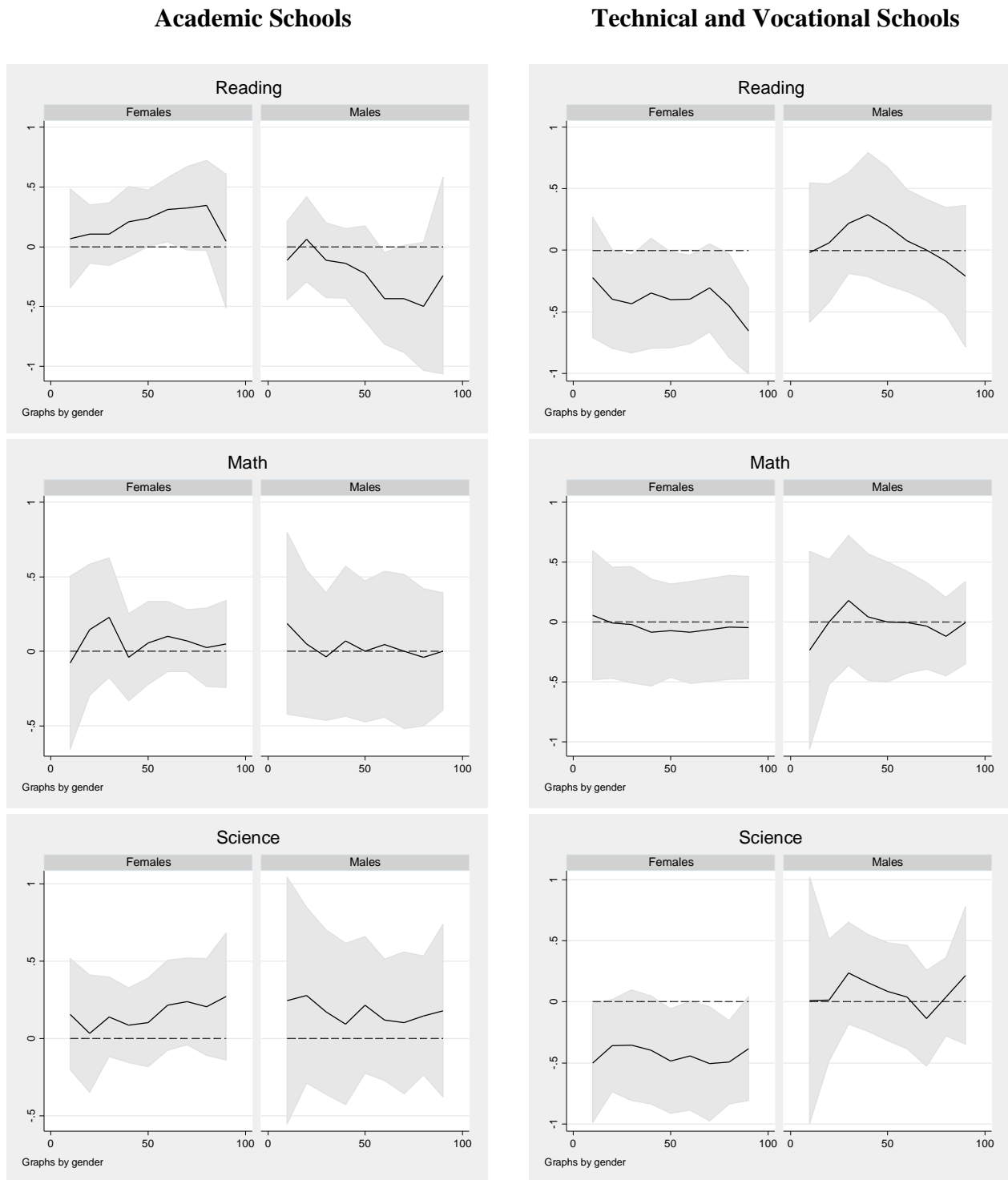|	Reading		Mathematics		Science	
	Males	Females	Males	Females	Males	Females
Effect	0.0384	-0.4194	-0.0351	-0.0039	-0.0055	-0.4184
	(0.1740)	(0.1490)	(0.1807)	(0.1559)	(0.1700)	(0.1501)
Baseline	415.6732	447.7098	481.5509	423.1282	459.6394	455.3818
	(10.8549)	(9.7333)	(16.4587)	(14.8501)	(13.6718)	(9.8758)

Notes. Results were obtained for groups defined by gender and curricular type (academic and technical or vocational), using the propensity score weighting estimator in Imbens (2004) using variables in Table 3. Baseline figures refer to reading, mathematical and scientific literacy in areas not affected by the reform (province of Trento). The effect is reported in points of standard deviation, after having standardized outcomes to have zero mean and unit variance. Standard errors, in parentheses, were computed via bootstrap using 200 replications, clustering on class.

Table A.2. Average effects on scores (sensitivity analysis - Main Sample).

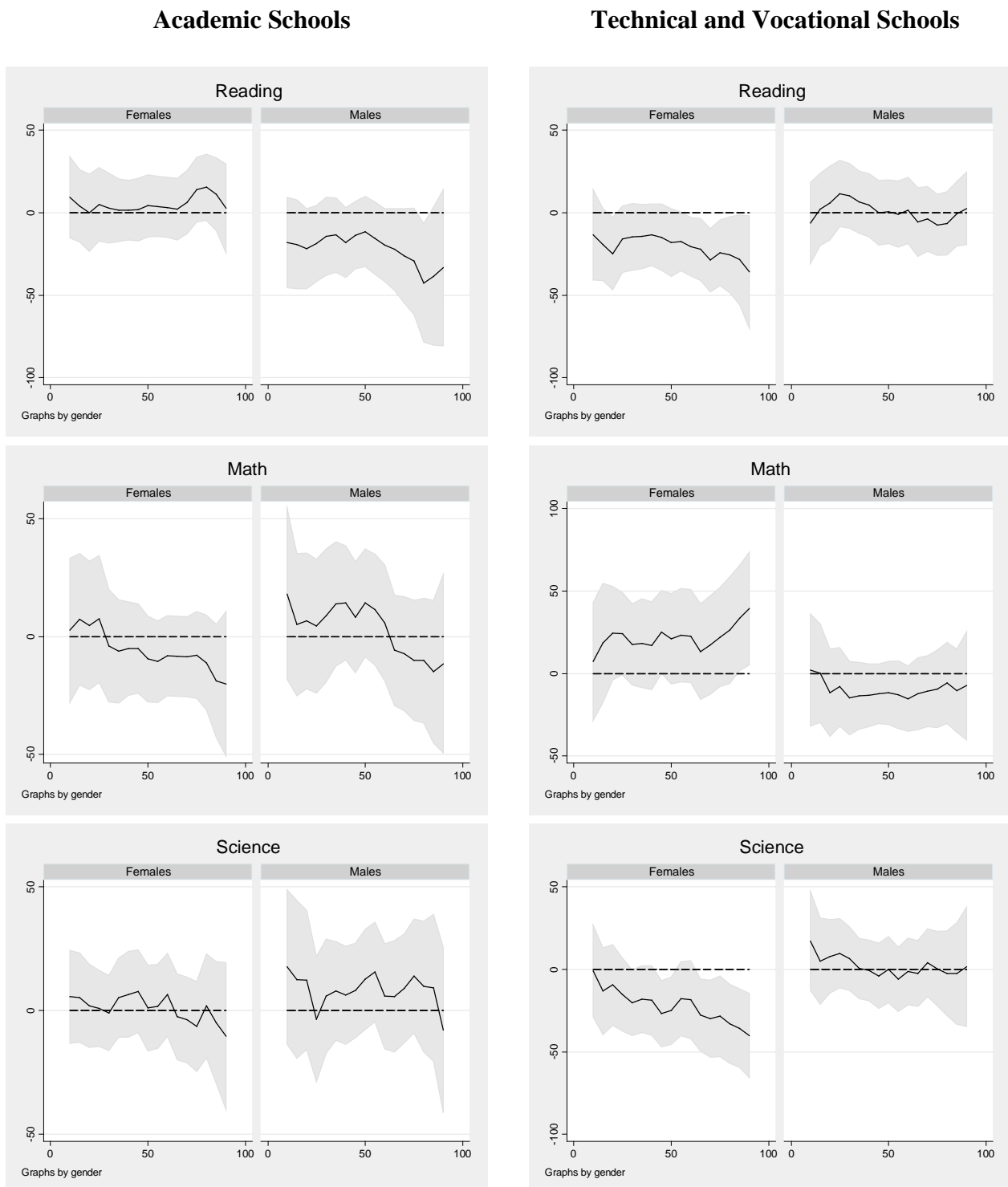|	**Reading**		**Mathematics**		**Science**	
	(1)	(2)	(3)	(4)	(5)	(6)
	Males	**Females**	**Males**	**Females**	**Males**	**Females**
Academic Schools						
	-0.3015	-0.0565	0.0230	-0.1061	0.1144	-0.0629
	(0.0848)	(0.1154)	(0.1071)	(0.1120)	(0.1074)	(0.0904)
Technical and Vocational Schools						
	0.0340	-0.2945	-0.0767	0.1099	0.004	-0.3107
	(0.1230)	(0.0841)	(0.1311)	(0.0904)	(0.1194)	(0.0891)

Notes. Results were obtained from separate regressions by gender and curricular type (academic and technical or vocational). All regressions include the variables in Table 3, and a proxy for student's ability obtained as explained in Section 6.2. The effect is reported in points of standard deviation, after having standardized outcomes to have zero mean and unit variance. Robust standard errors, clustered on class, are reported in parentheses.

Figure A.1. Quantile treatment effects on scores (sensitivity analysis - Main Sample).

**Academic Schools**   **Technical and Vocational Schools**

Notes. Reported are estimates of quantile treatment effects for reading, mathematical and scientific literacy. Results were obtained from separate regressions by gender and curricular type (academic and technical or vocational), using the method developed in Firpo (2010). Bootstrap standard errors based on 200 replications and clustered on class are used to compute 95% confidence intervals.

Figure A.2. Quantile treatment effects on scores (sensitivity analysis – Main Sample).
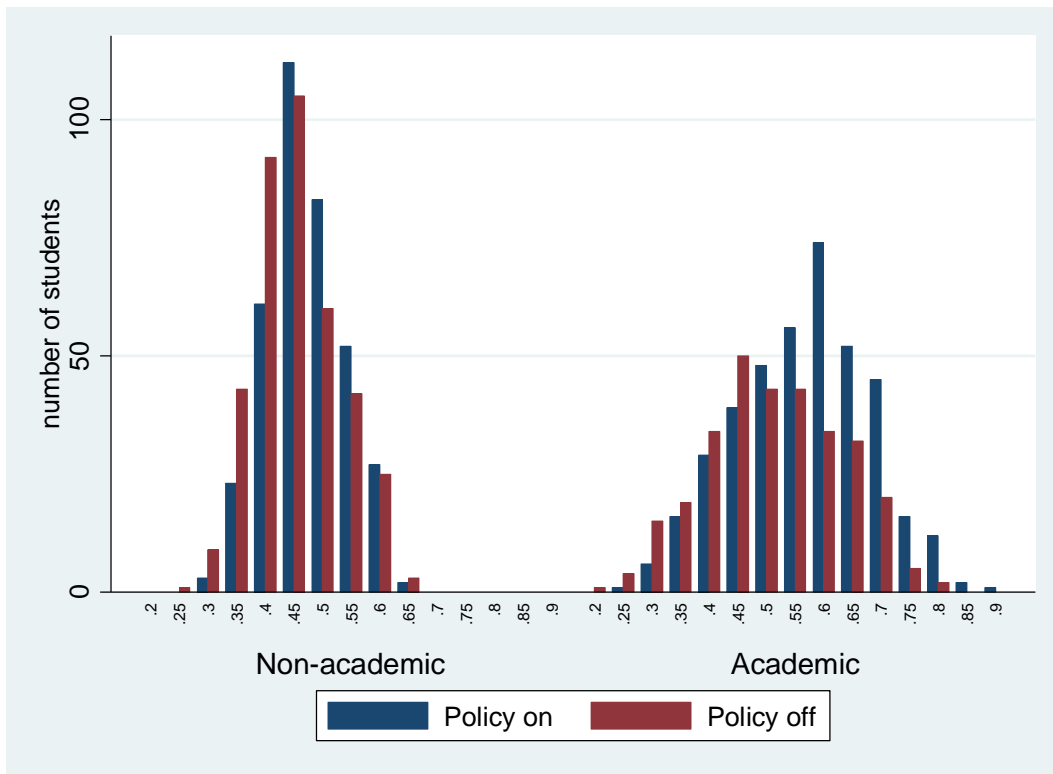

Notes. Reported are estimates of quantile regressions for reading, mathematical and scientific literacy. Results were obtained from separate regressions by gender and curricular type (academic and technical or vocational). All regressions include the variables in Table 3, and a proxy for student's ability obtained as explained in Section 6.2. Bootstrap standard errors based on 200 replications and clustered on class are used to compute 95% confidence intervals.

Figure A.3. Percentage of university graduates by cohort of birth (1994 reform).
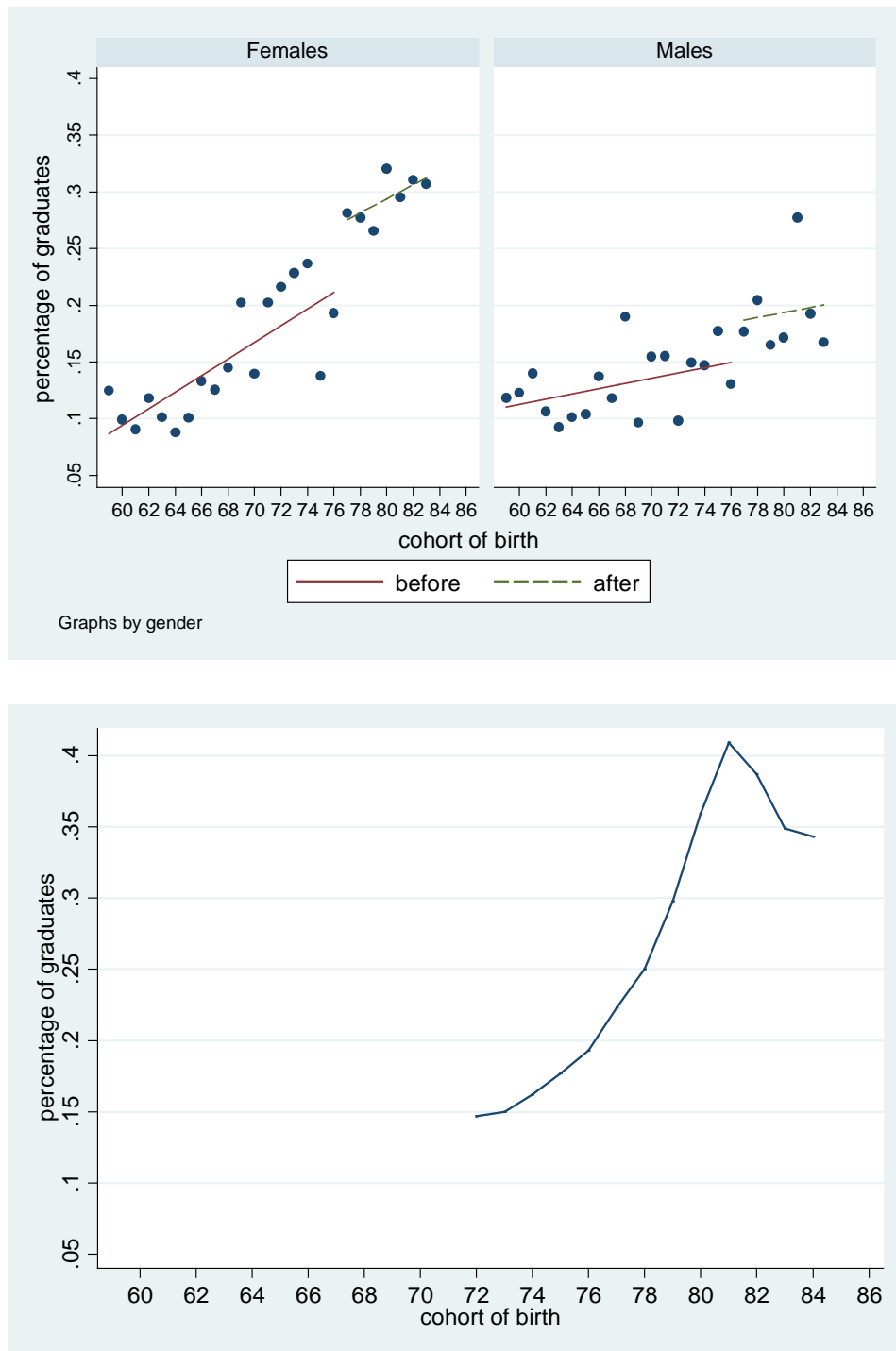
Notes. Reported is the percentage of university graduates by cohort of birth computed in 2010 from the Bank of Italy Household Survey on Income and Wealth (top panel), and using figures published from ISTAT (Italian National Bureau of Statistics: bottom panel). Cohorts born before 1976 are not affected by the 1994/95 reform of the upper secondary school system (see Section 6.4 for details).

Figure A.4. Propensity score distributions by curricular type.

Notes. Reported are predictions from a probit regression of "treatment" dummy on student characteristics in Table 3. Separate regressions are estimated by curricular type (academic or technical and vocational schools; the latter group is labelled "non-academic' here).