Time Spent at School and Inequality in Students' Learning Outcomes*

Giulia Bovini† Marta De Philippis‡ Paolo Sestito§

November 2016

Abstract

This paper analyzes the effect of increasing the amount of time spent at school on the dispersion and distribution of students' performance. It also evaluates whether spending more time at school instead of at home mitigates the importance of family inputs and therefore helps narrowing achievement gaps by socio-economic background. It studies the case of Italy, where two instructional schemes - that entail a different amount of time spent at school - coexist in primary schools, namely the *tempo pieno* scheme (TP, 40 hours per week) and the *tempo normale* scheme (TN, 24 to 30 hours per week). Identification comes from plausibly supply-driven variation in the share of TP classes offered within a given school-grade over subsequent cohorts of second and fifth graders. While the effect of TP is virtually null and relatively constant across deciles of the school-grade distribution of reading test scores, it is positive on mathematics and stronger at the bottom than at the top of the distribution. Switching from having no TP classes to having only TP classes would raise the first (ninth) decile of mathematics test scores by 3.6% (0.8%). A modest mitigating effect on achievement gaps by socio-economic background emerges, although coefficients are not precisely estimated. Effects are stronger in fifth than in second grade. Based on student questionnaires and Time Use Surveys, we document that that these findings are likely not driven by a change in total time devoted to instruction or leisure, but rather by changes in how time is allocated between autonomous study at home and supervised study at school.

*We thank Esteban Aucejo, Andrea Ichino, Nicola Pierri, Michele Pellizzari, Steve Pischke, Lucia Rizzica, Marco Tonello, Eliana Viviano and seminar participants at the Bank of Italy lunch seminar for very useful comments. We thank the INVALSI statistical office for their support as well as Claudio Quintano for providing us information on the algorithm to correct test scores from cheating. Finally we are indebted to Gianna Barbieri, Angela Iadecola and Antonio Martino at the Ministry of Education for providing us the data on school applications. The views expressed in this article are those of the authors alone and do not necessarily reflect the official views of the Bank of Italy.

†LSE and CEP

‡Bank of Italy and CEP

§Bank of Italy

1 Introduction

Understanding how the different inputs that enter the education production function affect students' average returns and attainment gaps is a core topic in the economics literature. The amount of time pupils spend at school is a costly, but relatively scalable school input, which displays substantial variation across OECD countries (OECD [2016b]). Yet, its causal effect on student achievement is relatively understudied compared to other education resources (e.g. class size). While most studies focus on changes in instruction time, time spent at school also encompasses time devoted to revising the curriculum, as well as recreational breaks and activities. Furthermore, despite possibly affecting students differently depending on the quality of opportunities available at home, the evidence on its effect on the dispersion and the distribution of achievement is scarce.

This paper studies how an increase in the amount of time spent at school - both in formative and recreational activities - affects the dispersion and distribution of pupils performance, as well as achievement gaps by socio-economic background. We exploit the coexistence in Italian primary schools of two instructional schemes that entail a different amount and organization of time spent at school, namely the *tempo normale/modulare* scheme (henceforth TN, where pupil spend at school typically 27 hours per week) and the *tempo pieno* scheme (henceforth TP, where pupils spend at school 40 hours per week). While under both schemes the same core content is covered, pupils on the latter typically have a longer school-day, as they have lunch and revise the curriculum at school, under the supervision of their teachers. More time spent at school may affect both the allocation of homework across school and home and the total time dedicated to instruction and leisure, depending on the substitutability or complementarity of time spent studying at school and at home. We investigate whether the increase brought about by the TP scheme in the amount of time devoted to study and revision at school - where the quality of inputs is relatively standardized and homogeneous - leads to a reduction in the amount of time devoted to these activities at home - where the quality of inputs varies greatly and family resources matter. Moreover, we evaluate how this affects the dispersion of learning outcomes and whether it helps providing a more level playing field for pupils from disadvantaged backgrounds.

Data on standardized test scores come from INVALSI.[1] Alongside test scores, INVALSI records a rich set of background information (e.g. citizenship, parental education

[1] Istituto Nazionale per la Valutazione del Sistema educativo di Istruzione e di Formazione.

and occupation, household resources at home), which is crucial for the study of achievements gaps. It also surveys students about their use of time outside school: combined with information from the 2008 - 2009 ISTAT Use of Time Survey, this allows to assess how the TP scheme affects total instruction and leisure time. The first way we use to identify the causal effect of interest exploits within-school-grade variation in the share of TP classes across subsequent cohorts of second and fifth graders, which comes from arguably supply-driven changes in the diffusion of TP schemes. As explained in Section 2, the supply of TP classes depends both on the school receiving enough demands for TP to activate such a class and on the availability of financial resources.[2] Our identifying assumption is that - once we control for observable characteristics and we allow for school-grade linear trends - the year-to-year variation in the share of TP classes within a given school-grade reflects idiosyncratic fluctuations in the amount of financial resources available or marginal changes in the demand for TP classes around the thresholds for class formation, whilst it is not systematically correlated with changes in the characteristics of students (and possibly teachers). We provide two pieces of supporting evidence. First, we use unique information provided by the Ministry of Education (MIUR) about applications for and actual enrollment in TP classes for the scholastic year 2014/2015 to show that unmet demand for TP schemes exists and to analyze where it is concentrated. Second, we perform a set of balancing tests to show that, within a given school-grade, changes in the share of TP classes over time appear not to be systematically associated to changes in the observable characteristics of pupils. Moreover, we plan to implement an alternative identification strategy, which is based on a regression discontinuity design that exploits jumps in the likelihood of activating a TP class generated by variations in the total number of applicants around the minimum class size threshold. Since test scores for the cohort of interest will become available shortly, Section 7 will present for the moment only the results from the first stage regression of the alternative identification strategy and the corresponding balancing tests.

There are four main findings. First, a larger share of TP classes has a virtually null impact on reading mean test scores, while it positively and significantly affects mathematics mean test scores. Second, while the effect on reading is relatively homogeneous across the distribution, the effect on mathematics is stronger at the bottom than at the top: switching from having no TP classes to having only TP classes would raise the first (ninth) decile of mathematics test scores by 3.6% (0.8%). This implies that the reduction in the inequality of learning outcomes is larger for mathematics. Third, there is evidence of a modest narrowing of achievement gaps by socio-economic background, although co-

[2] The TP scheme requires contracting more teaching hours, as well as offering the lunch service in a dedicated area.

efficients are not precisely estimated. All these effects are stronger in fifth than in second grade, suggesting the presence of a cumulative pattern. Fourth, combining information about pupils' use of time from INVALSI and from ISTAT Time Use Surveys, we show that what it is likely to drive these results is not a change in total time devoted to instruction or leisure, but rather changes in how time is allocated between autonomous study at home and supervised study at school.

This paper speaks to the literature that studies the effect of time spent at school on achievement. The early literature mainly focuses on the U.S. and exploits variation either across and within states (Rizzuto and Wachtel [1980], Card and Krueger [1992]) or across schools (Grogger [1996], Eide and Showalter [1998]). These studies document modestly positive to insignificant effects on years of completed education and earnings, but the leveraged variation is small, which may affect the precision of the estimates. Lee and Barro [2001] and Wößmann [2003] provide cross-country evidence using internationally comparable test scores and report similar findings. However, due to data limitations, both studies cannot include countries fixed effects to control for unobserved heterogeneity across countries. On the other hand, Lavy [2015], who analyzes 2006 PISA data within a student fixed-effect framework that exploits variation in instructional time across different subjects, finds a positive and significant effect, larger in schools with better institutions.

The recent literature seeks to improve identification and exploits instead plausibly exogenous variation in the number or length of schooldays stemming from natural experiments. One strand is based on weather-induced natural experiments, as it relies on changes in the number of schooldays prior to standardized tests because of unplanned school closures due to bad weather conditions (Marcotte [2007], Marcotte and Hemelt [2008], Hansen [2011], Goodman [2014]). These studies typically find positive effects. The other strand is based on policy-induced natural experiments, as it relies on reforms of the length of the term and/or the schoolday and on changes in term or test dates. Pischke [2007] studies the consequences of shortening the 1966-67 German school year, while Parinduri [2014] studies the consequences of lengthening the 1978-1979 Indonesian school year. They both show that longer school terms improve various measures of educational attainment. While the former does not find any long-lasting effect on earnings and employment, the latter reports a positive effect on these outcomes as well. Agüero and Beleche [2013] and Aucejo and Romano [2014] rely instead on variation in the number of schooldays prior to standardized tests because of changes in starting term dates and/or test dates, in Mexico and in the U.S. respectively. They all find beneficial effects as well, though the latter argue that the impact of reducing absenteeism is much larger than the impact of adding days of instruction. Closer to us, Bellei [2009] and Lavy [2012] analyse reforms that increase daily instructional time rather than the length of the term, in Chile and Israel respectively. They both conclude that longer schooldays boost achievement. Mariani et al. [2012] analyse the effect of TP schemes on learning outcomes as well, focusing on effects on average test scores.

This paper contributes to the literature in several ways. First, most of existing works rely on either small and unexpected or planned but short-lived variations in the length of the school-year, whereas we analyze a setting that features a substantial, known in advance and permanent difference in the length of the school day between different schemes. Given evidence of non-linearities in the productivity of time spent at school (OECD [2016a]), a very large change in the way students allocate their time between school and home may have implications that are not easy to grasp in a setting where the change in time spent at school is only marginal. Moreover, while unplanned changes leave little room to adjust the curriculum accordingly, teachers of TP classes know it in advance and can devise a reasoned teaching and pedagogical plan.

Furthermore, this paper is among the few works studying the effect of increasing the amount of time spent at school on the entire distribution of test scores, on its dispersion and on achievement gaps by socio-economic background. To our knowledge, the only paper that explicitly looks at heterogeneous effects across different deciles of the distribution is Huebener et al. [2016]. However, their setting is very different: they analyze the effect of an increase in weekly instruction time in German secondary academic track schools, which translated into new content being taught during additional hours. They provide evidence of a widening, rather than shrinking, gap between low- and high-performing students. We instead study much younger pupils and a scheme under which additional classroom time is mostly devoted to revising the curriculum. We are also among the few to provide evidence about how increasing time spent at school affects the use of time outside school.

Finally, we collect detailed and previously unexploited information on the intensity and the characteristics of the demand for longer school days. At this stage, we use this data to analyze the existence and the pattern of unmet demand for TP schemes. When we will be able to match application data to achievement data, we will also use a different identification strategy, as explained before.

The remainder of the paper is organized as follows. Section 2 provides an overview of the institutional framework. Section 3 describes the data. Section 4 lays out our two alternative estimation strategies. Section 5 displays the main results from the first identification strategy. Section 6 provides evidence about the use of time outside school. Section 7 discusses our alternative identification strategy we will fully implement once more data become available. Finally, Section 8 concludes.

2 Institutional setting

The Italian school system is organized around three cycles: primary schooling (grades 1-5, from age 6 to age 11), lower secondary schooling (grades 6-8, from age 11 to age 14) and upper secondary schooling (grades 9-13, from age 14 to age 19). Primary and lower secondary schools are single-tracked, while upper secondary schools offer three tracks, namely academic, technical and vocational. Students are assigned to a specific class for the entire length of any educational cycle and classmates are the same for all subjects. Education is compulsory up to age 16.

We focus on primary school because it is the cycle where the difference in the length of the school day across instructional schemes is particularly large. Indeed, since the end of the 1960s, pupils in Italian primary schools can apply to two different instructional schemes: the TN and the TP.[3] The choice is made at the beginning of primary school, normally at the age of 6, and the classes formed according to one of the two schemes are later maintained until a given cohort reaches the fifth grade.[4] Under TN schemes lectures cover no less than 24 hours per week (usually 27), distributed across five to six school days per week, typically only in the morning. Under TP schemes pupils spend 40 hours per week at school, split across five days per week, normally from 8.30 a.m to 4.30 p.m. While the same core content is covered under both schemes, students logging longer school days have lunch and revise the curriculum at school, under the supervision of their teachers and in at least partial replacement of homework. Other diverse extra-curricular activities may take place during additional school hours, which may entail a more intensive use of laboratories and may aim also to help pupils to socialize among themselves and build relational skills. Since the late 1980s, pupils in TN classes are assigned to three main teachers every two classes, while pupils in TP classes are assigned to two main teachers. All classes may have the additional support by some specialized

[3]The TP scheme has been introduced in Central Italy at the end of '60s (and formally established by law 820/1971).

[4]Over time a class may get bigger or smaller both because some students fail to pass at the successive grade (in our data, 3% of second grade students and 5% of fifth grade students have a delay in their study path) and because of mobility from one school to another (in the scholastic year 2014/2015 the share of transferring students was 1.7% for second graders and 1% for fifth graders (source: http://cercalatuascuola.istruzione.it/cercalatuascuola/)).

teachers for specific subjects (religion, foreign languages etc.).[5,6]

Hence, TN and TP classes differ not only in the amount and organization of time spent at school, but also in terms of the number of teachers and, possibly, in terms of teaching practices. Since we have no information about teachers in our dataset, in principle what we are able to estimate is the effect of attending TP classes relative to TN classes, comprehensive of the pedagogical differences between these two modules. However, while the jump from having only one teacher (as it was the case in TN classes before the end of the 1980s and is today in the few TN classes with 24 hours of instruction) to having more than one teacher may have some pedagogical implications – as the single teacher often acts as a very strong reference person for pupils –, the passage from two to three teachers is unlikely to have major consequences. Moreover, we will explore whether teaching practices, as well as preparation to INVALSI tests, differ across TN and TP schemes, since we are in the process of matching achievement data with questionnaires administered to a random sample of teachers. Therefore, we argue that the main difference between TN and TP schemes is the amount of time spent at school and the fact that TP schemes generally make pupils revise at school, which typically at least partially replaces homework.

Crucial for our identification strategy, which exploits variation within school-grade in the share of TP classes over time, is a discussion of the reasons why such changes may occur. In the winter prior to the start of primary school, parents express their preference for the school and the instruction scheme. The supply of TP classes does not always fully satisfy the demand for them. On the one hand, the law prescribes a lower bound to class size. If the demand for longer school days in a given school is insufficient, the school does not activate such a class.[7] On the other hand, as the law also indicates an upper bound to class size and the provision of TP classes is more costly, financial resources may be insufficient to accommodate all demands for TP instruction.[8]

[5]Before 2009, i.e. before the implementation of the so called "Riforma Gelmini" (laws 133/2008 and 196/2008), the two TP teachers were also concurrently present for at least two hours per week. We control for differences associated to the Gelmini reform by including a dummy that takes value 1 if a given cohort is exposed to the reform (which corresponds to only one cohort in our sample, since the reform has not affected classes that were already formed before it took place).

[6]Notice that while the Gelmini reform has introduced the possibility of having a single teacher and 24 hours of instruction, the take up of this option has been extremely low, 0.5% according to Battistin and Schüller [2013].

[7]The Gelmini reform (laws 133/2008 and 196/2008) has increased the minimum number of students in a class for school located in *comuni non montani* from 10 to 15, starting from first grade classes in the scholastic year 2009-2010. For schools located in *comuni montani* the minimum number is still 10. For both types of municipalities, the lower bound can be adjusted downward up to its 10%.

[8]Following the Gelmini reform, the maximum number of students in a class is 26. Such upper bound may be adjusted upward up to its 10%. As mentioned in Section 1 TP classes require additional financial resources, because of the need of contracting teachers for more hours and setting up a canteen for the lunch break that complies with legal standards.

Within a given school, if the demand for longer school days exceeds supply, the head teacher can ask for the activation of a new TP class. Schools, however, are very often constrained in activating such new classes by lack of public funds. In case of over-demand for TP, the school decides which students to allocate to the TP schemes[9], given the available resources[10] and infrastructures, based on criteria that often include the proximity to the place of residence, own or siblings' disability, siblings' enrollment in the same school, parents' employment and income. Weights of different criteria may vary greatly across schools, but they are usually constant over time.

3 Data

The paper combines data from three sources. Data about achievement come from nationwide, standardized tests that assess reading and mathematical skills. The tests are administered by INVALSI to the entire population of second and fifth graders (aged 7 and 11 respectively) at the end of the scholastic year, in two separate days during the month of May. Schools are required to provide information about the demographic and background characteristics of their students, which include among others age, gender, citizenship, regularity of studies (i.e. whether the pupil is ahead or behind her grade level), previous attendance of nursery and/or kindergarten, parental education and occupation. Schools also report class-level information, namely the number of students enrolled and the time schedule. Based on this, we flag a class as offering a TP scheme if the modal reported time schedule is 40 hours per week.[11] Fifth graders are also required to complete a questionnaire about several dimensions of their life inside and outside school. Our final panel dataset includes second and fifth graders, and follows schools[12] over the scholastic years from 2010-2011 to 2014-2015.[13]

[9]Since 50% of schools in Italy have only one class, in these cases the choice is about whether to admit a student or not.

[10]The availability of financial resources mainly depends on national funding, that covers roughly two third of total expenses in primary and secondary education. These resources are generally allocated in an homogeneous way along the country. The remaining share of funding, financed by local authorities, displays more variability.

[11]Within-class variation in the reported scheme is infrequent. It likely indicates either the presence of students with special needs who benefit from a customized instructional plan or it stems from a reporting error.

[12]By schools we mean the physical buildings (*plessi scolastici*). In some cases multiple schools are grouped into the same institution, managed by the same head teacher. However, the institution identifier is not consistent over time and cannot be used as a panel dimension.

[13]Up to the scholastic year 2008-2009 participation to the tests was voluntary. Participation to the tests has been compulsory since the scholastic year 2009-2010, but the panel structure of the dataset that allows to observe each school over subsequent cohorts of tested students only starts from the scholastic year 2010-2011.

The INVALSI dataset is comparable to education datasets available in other countries, but it has two important advantages that are precious for our analysis. First, the presence of richer information on household characteristics: while U.S. datasets typically only report eligibility for free or reduced price meals, INVALSI records information on parents' education and occupation, as well as on the number of books available in the household and the language prevalently spoken at home. This is crucial to asses whether increasing time spent at school helps narrowing achievement gaps by parental background. Second, the presence of students' questionnaires that investigate the way children spend their time after school. This information allows to understand whether and to what extent time spent at school crowds out study or leisure time outside school.

The dataset, however, has also some limitations. First, it contains no information about teachers. This precludes checking whether cohort-to-cohort changes in the share of TP classes within a school-grade are not systematically associated with changes in the composition and characteristics of the teaching staff, as our estimation strategy assumes. Moreover, INVALSI reports evidence of episodes of test manipulation, especially in primary schools and in some Southern regions.[14] Because of this, we use test scores corrected for a cheating factor computed at the class level since the 2011-2012 wave of the tests and provided by INVALSI.[15] This implies that the 2010-2011 wave is excluded when estimating regressions with scores as the dependent variables, but is used to provide descriptive statistics in Section 4.3 and to investigate students' use of time outside school in Section 6. Second, each year INVALSI randomly selects a sample of schools and sends external invigilators in one or two classes to supervise the administration of the tests and the transcription of students' answers into machine-readable answer sheets.[16] We include in all regression specifications a dummy for whether an external monitor was present in at least one class of a given school. Finally, in the Appendix we replicate the analysis using raw scores. The main findings are confirmed.

We also have access to a unique dataset on applications to primary schools provided by the Ministry of Education (MIUR). For each pupil starting primary school in the scholastic year 2014-2015, we observe the school and the time schedule (TN or TP)

[14]Tests are proctored by the school staff, although teachers are not supposed to administer the tests to grades they teach. Proctors have also to copy students' original responses onto machine-readable answer sheets that are then sent to INVALSI. Manipulation may thus occur during transcription. See Angrist et al. [2014] and Paccagnella and Sestito [2014] for related work.

[15]See Quintano et al. [2009] for a detailed explanation of the procedure used to compute the class-level cheating factor. The subsequent changes in the procedure are described in Falzetti et al. [2015]. An alternative procedure has been proposed by Pereda Fernández [2015]: the main stylized facts – and in particular the North–South difference do not differ very much.

[16]External invigilators are selected from a list of mainly retired teachers and school principals who have not worked in the town or the school they are assigned to for at least two years.

preferred by parents in their application form, submitted prior to the start of the scholastic year, in January and February. We also know the school and the time schedule (TN or TP) each students ends up being enrolled to in September, when the scholastic year starts. Pupils who start primary school in 2014/2015 take INVALSI tests for the first time in 2015/2016, when they reach the second grade. At the time of writing we are in the process of linking application data to achievement data. At this stage, data about applications and actual enrollment only provide some descriptive evidence about the intensity of total and unmet demand for TP schemes. In Section 7 we set out alternative identification strategies that we plan to explore once we will be able to link such data.

Finally, the 2008-2009 wave of the Time Use Survey administered by ISTAT is used to provide further information about differences in use of time between pupils attending TP and TN schemes.

4 Empirical strategy

As described in Section 2, the estimation of the causal effect of time spent at school on achievement is challenging because of the mechanism that assigns pupils and teachers to schools and to classes within schools. Parents self-select into the preferred school and the preferred scheme within a school depending on preferences that may be driven by characteristics not fully observable and possibly correlated with students' learning outcomes. School headteachers manage excesses of demand for a given scheme, if any, according to a set of criteria that are not observable either. Moreover, they may match teachers with specific pedagogical practices to different schemes. As a result, pupils (and possibly teachers) in TN and TP classes are likely to differ along unobservable dimensions that affect both the propensity to choose or be assigned to TP and test performance. Hence, the difference in achievement between students enrolled in TN and TP classes would not reflect the causal effect of interest.

In order to mitigate these selection problems, we propose two alternative identification strategies. The first relies on within school-grade variation in the share of TP classes over time and compares cohorts of students differently exposed to the TP option. Results are discussed in Section 5. The second identifies the effect through a regression discontinuity design based on the jump in the probability of activating a TP class according to whether the total number of TP applicants is above the minimum class size threshold. At the moment, we can only present in Section 7 the results from the first stage regression of the second identification strategy because we are still in the process of linking MIUR application data with INVALSI test scores, which will allow us to estimate the second stage of our fuzzy RDD design.

4.1 First identification strategy: within school-grade variation

This strategy deals with selection between schools by including school-grade fixed effects in our regression specification. Second, it deals with selection into classes within schools that offer both TN and TP schemes by aggregating information at the school-grade level. Third, to address the possibility that there exist time-varying unobserved factors, our regression specifications include a set of school-grade linear time trends, which control for linear changes in unobservable characteristics. Identification then comes from changes in the share of TP classes over time, across subsequent cohorts of second and fifth graders, in deviation from the school-grade long-run trend. This estimation strategy has been pioneered by Hoxby [2000] and used in several other studies, including Mariani et al. [2012] and Lavy et al. [2012]. The baseline regression specification therefore reads:

$$y_{sgt} = \delta ShareTP_{sgt} + X'_{sgt}\gamma + \mu_{sg} + \theta_{sg}t + \eta_t + \varepsilon_{sgt} \quad (1)$$

where s indexes the school, g indexes the grade and t indexes time. y_{sgt} is the (log) moment of interest of the distribution of test scores in school s, grade g and year t; $ShareTP_{sgt}$ is the share of TP classes, X'_{sgt} is a set of time-varying controls at the school-grade level;[17] μ_{sg} is a set of school-grade fixed effects; $\theta_{sg}t$ are school-grade linear time trends; η_t is a set of year fixed effects; ε_{sgt} is the error component.

The inclusion of school-grade fixed effects and trends comes at the cost of greatly reducing the variability in the share of TP schemes that can be exploited for identification, as most of the variation is cross-sectional rather than longitudinal. The year-to-year change in the share of TP classes is null for 88% of school-grades. An increase in the share of TP classes is witnessed in 8% of school-grades, while a decrease in the remaining 4%. The identifying assumption underlying our empirical strategy is that, conditional on observables and allowing for school-grade linear time trends, the variation over time in the share of TP classes within a given school-grade reflects plausibly exogenous variation in the supply of TP classes due to fluctuations in financial resources or marginal changes in the demand for TP across the thresholds prescribed by class size laws, whilst not being systematically correlated with characteristics of students (and possibly teachers).

[17] In particular the share of male students, the share of students who are Italian citizens, the share of regular students, as well as the share of students with mothers and fathers holding at least the high-school diploma, along with the share of students with missing values for each variable. Due to the large share of missing values (29.84%), the share of pupils who attended the nursery, although available, is not included among the covariates. Information about parental occupation is not used as a control, because - especially for mothers - it can be an outcome of the treatment themselves. Controls also include the school size and its square, the number of classes, a dummy for whether an external invigilator proctored the test in at least one class, a dummy for whether the Gelmini reform was in place.

Finally, we impose some restrictions to the sample for the analysis. First, we drop schools without an identifier that tracks them over time, which account for less than 0.004% of observations. Second, we discard school-grade-years where the number of students present the day of the tests is larger than the number of students officially enrolled, as this may reflect recording errors. This results in losing less than 0.001% of observations. Third, we drop school-grade-years where the instructional scheme is not reported for at least one class, which account for less than 0.03% of observations. Fourth, we discard school-grade-years where the number of students in at least one class is lower than the minimum set by the law. This accounts for roughly 10% of observations and is a customary restriction when analyzing primary schools data, because it likely eliminates schools in less densely populated and remote areas - typically mountainous communities or small islands - which may also adopt different pedagogical practices and schemes, making them not comparable.[18]

4.2 Alternative identification strategy: fuzzy regression discontinuity design

The alternative identification strategy leverages variation in the schools' probability of offering TP classes that depends exclusively on class size rules. In Italy the law prescribes a lower bound to class size: 15 pupils in the so-called *comuni non montani* (CNM)[19] and 10 pupils in the so-called *comuni montani* (CM). In both cases a 10% tolerance buffer applies. Therefore, there exists a discontinuity in the probability of offering at least one TP class at the start of the school year in September depending on whether more than 13 (or 9 in CM) students expressed a preference for TP schemes when applying to primary school in February.[20]

As stated before, class size rules are not the only source of variation in the schools' probability of offering TP classes: schools may also be constrained by the absence of enough teachers or of the appropriate infrastructure, for instance. Moreover, between the moment when headteachers review applications in February and the moment when classes are formed in September, adjustments to parental preferences may occur. We therefore intend to adopt a fuzzy regression discontinuity design: we use the discontinuous jump

[18] Angrist et al. [2014] and Ballatore et al. [2015] make a similar restriction.

[19] It literally means *cities not in a mountain area*, the distinction deriving from the fact that cities denominated as *in a mountain area* enjoy some exceptions to class size rules because they are considered communities more difficult to reach.

[20] The discontinuity will also apply, for large enough schools, to the possibility of activating two or more classes, depending on the number of applications to TP. For now, we look for simplicity only at the first discontinuity, which is the one where most of the variation comes. We will then exploit all discontinuities in our regression specifications.

in the probability of offering TP classes that depends on class size rules as an instrument for the provision of TP classes in a given school-grade. Of course other characteristics of the school and of the cohort may change with the number of applications for TP, but they are very unlikely to change in a discontinuous way and exactly around the minimum class size threshold. We will therefore estimate the following equation:

$$y_s = \gamma TP_s + f(D_s^{TP}) + u_s \tag{2}$$

where we instrument the variable TP_s exploiting the discontinuity around theminimum class size threshold:

$$TP_s = \beta I(D_s^{TP} \geq \bar{D}_s) + f(D_s^{TP}) + \epsilon_s \tag{3}$$

where y_s is the outcome variable (a given moment of the distribution of students' test scores in grade g in school s for the 2014 , for which we have information on applications); TP_s is a dummy equal to one if school s offers at least one TP class; D_s^{TP} is the number of applications to TP received in school s; $\bar{D}_s$ is the lower bound to class size that applies in school s and $f(D_s^{TP})$ is a polynomial of the number of TP applications received by the school. Finally, ϵ_s is an error term.

We therefore compare schools receiving a number of applications for TP schemes within a small interval around the cut-off point, the underlying idea being that these schools, and their students, will be identical along all characteristics with the exception of eligibility for an extra (TP) class. At the time of writing, we are in the process of matching MIUR application data with INVALSI test scores. Therefore, we can only estimate the first stage of this identification strategy, i.e. equation (3) only.

4.3 Descriptive evidence

In Section 2 we discussed two reasons why supply of TP classes may not fully satisfy demand: lower bounds on class size and availability of financial resources. MIUR data about applications to primary schools and enrollment in first grade for the scholastic year 2014-2015 provide insightful evidence about this. The map in Figure 1 shows the intensity of parental demand for TP classes, expressed as a fraction of total applications, across provinces. A remarkable geographical heterogeneity emerges: demand is much stronger in the North than in the South (the region of Basilicata being an exception). Parents preferring TP schemes range from as many as 85% in the province of Milan, the capital of the Northern region of Lombardia, to as few as 2% in the province of Ragusa, located in the Southern region of Sicily. The map in Figure 2 displays the share of students actually enrolled in TP classes at the provincial level. The heterogeneous intensity of demand for TP across provinces translates into heterogeneous diffusion of TP schemes. The map in Figure 3 reports instead the share of applications for TP classes that is not satisfied (i.e. parents apply for the TP scheme, but their children ends up being enrolled to a TN scheme). It documents that supply of TP does not always fully satisfy the demand. The intensity of unmet demand tends to be higher in provinces with either very low or high intensity of demand, suggesting that both lower bounds to class size (in areas with weak demand) and insufficient resources (in areas with strong demand) may play a role in the rationing of supply.

Table 1 shows school- and student-level descriptive statistics for the main sample from INVALSI, split by grade. Primary schools are relatively small in size, catering in a given grade less than 35 students on average. The average class size is around 17, slightly larger for second graders due to the increase in lower and upper bounds to class size introduced by the Gelmini reform. The share of TP classes is 0.33 in second grade and 0.29 in fifth grade, suggesting a slow increase of the diffusion of TP over time. Students with the Italian citizenship constitute 90% of the students and only 3% to 5% of students are either ahead or behind their grade level. Almost 10% of pupils speak prevalently a foreign language at home. 51% (60%) of fathers (mothers) hold at least a high-school diploma.

Table 2 documents further differences among TN classes and TP classes. Panel I shows that TP classes are more likely to be located in densely populated municipalities and in provincial capitals. Consistently with a greater diffusion in the Centre-Northern Italy and in urban areas, TP classes are more widespread in areas with more active local labor markets, where the unemployment rate and the female non-participation rate are lower. Panel II refers to school and student characteristics. Average class size is bigger for TP schemes than for TN schemes. This reflects the fact that TP classes are more diffused in urban contexts, but also it probably reflects the fact that there exists some rationing of TP. Relevant differences in terms of pupils demographic and background characteristics emerge as well. TP classes cater for a larger share of non-native students. Pupils who attend TP classes are much more likely to having attended the nursery and having working mothers. They are also slightly more likely to have parents who hold at least the high-school diploma. The ESCS indicator - an indicator of socio-economic status that is available for fifth graders only and summarizes information about parental education, occupation and wealth - is higher. As pointed out in Mariani et al. [2012], a simple regression of the probability of attending TP classes on such ESCS indicator and its square reveals that the relationship is inversely U-shaped. This means that the propensity to demand TP schemes is lower on the left tail (where parents may not benefit from time subtracted to child-care) and on the right tail (where parents may afford other and more customized types of child-care services) of the ESCS distribution.[21] Overall, facts in Panel II fit nicely with the observation that TP is more diffused in urban areas and in regions where the provision of child-care services is higher and women are more likely to work.

Panel III shows reading and mathematics test scores, which are standardized by subject, grade and year to have mean 0 and standard deviation 1. While students enrolled in TP classes perform worse on both subjects when looking at raw test scores, they perform better when looking at test scores corrected for cheating. This is likely explained by the fact that TP schemes are less diffused in areas where cheating correction factors tend to be larger.

5 Results exploiting within school-grade variation

5.1 Checks to the identification strategy

The identifying assumption underlying this estimation strategy is that, once selection into schools and into classes within schools is properly accounted for, cohort-to-cohort changes in the share of TP classes within a given school-grade are not systematically associated to unobserved changes in the characteristics of students or teachers. They rather reflect fluctuations in the resources available to provide longer school days or marginal changes in demand for TP classes around thresholds relevant for class size rules.

Table 3 shows the results of a set of balancing tests on observable school and students characteristics, namely: enrollment; absenteeism on the days of the tests; the share of male students; the share of students who are Italian citizens; the share of regular students; the share of students whose mothers and fathers have at least the high-school diploma; the share of missing values for each of these variables. If observable characteristics are correlated with unobservable ones, balancing tests may provide some intuition on the amount of selection on unobservables that our estimates capture. When not including school-grade fixed effects (column 1), a statistically significant association between the share of TP classes and these characteristics exists. The correlation weakens when school-grade fixed effects are introduced (column 2) and is almost completely wiped away when school-grade linear trends are added as well (column 3). A weak association survives only with enrollment and the share of students who do not report their citizenship.

[21] The results are not shown for brevity, but are available upon request.

Overall, the introduction of school-grade fixed effects and trends eliminates most of the association between the share of TP classes and characteristics of pupils in a given school-grade. We take this as a signal that the association with their unobserved characteristics is likely to be non significant either.[22] Nothing can instead be said on teachers as there is no available information.

5.2 The effect of attending the TP scheme

Tables 4 and 5 show the effect of TP on the deciles of the test scores distribution. The tables report the coefficients relative to the share of TP classes coming from a set of ten regressions as specified by equation (1), where the dependent variables are the (log) values of the deciles of the school-grade level distribution of reading and mathematics test scores, respectively. Figures 4 and 5 plot those coefficients, along with 95% confidence intervals. Pooling second and fifth grades together, it appears that the effect of TP on reading test score is virtually null and non significant across all deciles of the distribution. When splitting grades, attending TP appears to have a slightly negative effect in second grade and a slightly positive effect in fifth grade, although coefficients are never significant. Conversely, the effect of TP on mathematics test scores is positive and significant across all deciles of the distribution. Moreover, it is stronger at the bottom than at the top: switching from having no TP classes to having only TP classes would raise the first decile of the school-grade distribution by roughly 3.6%, while it would raise the ninth decile by 0.8%. Specifications split by grade reveal that the heterogeneous effect across the distribution is larger in fifth grade.

Table 6 highlights the effect of TP on the log of mean and median test scores. Consistently with the findings coming from tables 4 and 5, attending TP schemes does not have a significant impact on the mean and the median of reading test scores. On the other hand, the effect on mathematics test scores is positive and significant: according to the pooled regression, switching from having no TP classes to having only TP classes would raise the average mathematics score in a given school-grade by roughly 2.2%. For both subjects, effects are more positive in fifth grade.

Table 7 focuses on measures of dispersion, namely the log of the test scores' standard deviation, the ratio between the third and the second quartile of the distribution (75-25 ratio) and the 90-10 ratio. A larger share of TP classes reduces inequality in learning

[22]We also run balancing tests on the differences in the characteristics of children under TP and TN in schools that offer both TP and TN classes. This specification checks that there are not large differences in the composition of TP and TN classes correlated with the increase in the share of TP classes. On average, we find that an increase in the share of TP classes is not correlated with changes in the difference in average socio-economic characteristics between students under TP and TN.

outcomes. While the compressing effect is small and not significant for reading, it is larger and often statistically significant for mathematics: for instance, switching from having no TP classes to having only TP classes would reduce the ratio between the ninth and the first decile of the distribution by roughly 2.8%. For both subjects, coefficients are larger for fifth grade. Overall, this is consistent with the TP having a stronger effect at the bottom than at the top of the distribution of mathematics test scores, especially in fifth grade.

Table 8 shows the effect of longer school days on (log) achievement gaps (expressed as ratios) by socio-economic background, along four dimensions: mother and father education (diploma/not diploma); citizenship of the students (Italian/foreign); language spoken prevalently at home (Italian or a regional dialect/foreign language).[23] Overall, TP schemes appear to have a small mitigating effect on achievement gaps, which is largest when grouping students according to the language prevalently spoken at home. However, most coefficients are not precisely estimated, also because of the smaller sample size.[24] Mitigating effects tend to be larger in fifth grade, consistently with findings about the effect of TP on dispersion and inequality of learning outcomes.

6 Use of time

As discussed in the introduction, varying the amount of time spent at school may induce changes in the amount of time devoted to study at school - where the quality of inputs is relatively standardized and homogeneous - and in the amount of time devoted to study at home - where the quality of inputs varies greatly and family resources matter. Depending on the substitutability or complementarity of these inputs, total time dedicated to instruction and to leisure can change as well. This is one channel through which attending TP schemes rather than TN schemes may affect the distribution and the dispersion of learning outcomes.

INVALSI questionnaires administered to fifth graders collect information about their use of time outside school.[25] Table 9 reports the frequency of homework and different recreational activities, splitting pupils according to the time scheme attended. On the one hand, TP students devote less time to homework: for example, the share of students

[23]The information about the language prevalently spoken at home comes from the student questionnaire. Hence, it is available only for fifth graders.

[24]For each background characteristic considered, the regression is performed only on school-grades catering for both types of students (e.g. both Italian and foreign students when studying achievement gaps by citizenship.)

[25]This information is collected in the 2010-2011, 2011-2012 and 2012-2013 waves of the questionnaire only.

reporting to do homework more than 5 times a week drops from 64% under the TN scheme to 24% under the TP scheme. Therefore, TP schemes entail an at least partial substitution between study at home and study at school. On the other hand, there are not sizable differences in the frequency of leisure activities (e.g. playing sports, reading books, playing with friends). This indicates that the total time devoted to leisure is not largely affected by the instructional scheme attended. Table 10 confirms these findings within a regression framework as the one outlined in (1), where the dependent variables now relate to time use. While the share of TP classes in a given school-grade is negatively and significantly associated to the frequency of homework, no significant effect on the frequency of recreational activities emerges.

The Use of Time Survey administered by ISTAT to a random sample of households in 2008-2009 allows to further investigate how strong is the substitution between study at home and study at school, and hence what happens to total time devoted to instruction.[26] Figure 6 confirms that students enrolled in TP schemes receive more classroom instruction and do less homework than students enrolled in TN schemes from Monday to Friday. On Saturdays, the former do not attend school, while the latter do, and the amount of homework is virtually the same. Overall, TP pupils devote to instructional activities (i.e. classroom instruction time and homework) roughly two hours more than TN pupils per week, which amounts to a relatively modest increase of 20 minutes a day, less that 5% of the total instructional time. This suggests that the substitution between revision at school and revision at home is close to being 1:1 and there are not sizable differences in the total amount of time devoted to instruction.

Finally, Figure 7 investigates whether pupils under TP or TN schemes differ in terms of help received from their parents while doing homework. It appears not, in relative terms, as parents assist children with homework roughly 40% of the time regardless of the instructional scheme. However, since the total time spent doing homework at home is much higher for TN students, in absolute terms the differences are large.

7 Results from the fuzzy regression discontinuity

We first show the results from our first stage regression of the fuzzy RDD design graphically. Figure 8 displays the pattern of the discontinuity of interest for *comuni non montani*

[26]The Use of Time Survey does not explicitly records the instructional scheme the student is enrolled to. Focusing on the sample of children who attend primary school and are interviewed from October to May and from Monday to Saturday - i.e. during the months and days when primary schools are open in Italy - we assign pupils to TP schemes if they report that the last class ends after 2p.m, whereas we assign pupils to non-TP schemes if they report that the last class ends before or at 2p.m or if they report to attend school on Saturdays.

and *comuni montanti*, respectively. As expected, our RDD is fuzzy: there are schools that manage to activate a TP class in September even if there were not enough students applying for TP in February, as well as schools that do not activate a TP class even if enough students asked for it, because of lack of funds for instance. Yet, we observe a sizable jump around the discontinuity.

Column 1 of Table 11 displays results from the first stage regression. The top panel includes the entire sample of applicants and fits a three degree polynomial of the running variable, i.e. the number of TP applicants to school s; the bottom panel includes instead only schools whose number of TP applicants is around the minimum class size threshold (below 25). Our estimates show that if the number of TP applications exceeds the minimum class size requirements, the likelihood of actually offering a TP class increases by about 10 percentage points. Note that for the strategy to be valid, the fact that data on applications may not fully capture the entire demand for TP schemes - if parents who know that there is no chance of their school activating a TP class do not ask for it - is not problematic, because identification relies on schools at the margin of offering or not a TP class.

Figure 9 as well as columns 2, 3 and 4 of Table 11 provide some checks to our identification strategy. In order for a RDD strategy to yield consistent estimates, the counterfactual distribution of the outcome variable must be smooth enough around the discontinuity of the running variable. This means that the average and the standard deviation of students' test scores in a school, in the absence of the discontinuity derived by minimum class size rules, must be smoothly related to the number of students applying for TP classes. A common way to assess the validity of this assumption is to check that other variables usually associated with the outcome of interest do not also vary discontinuously at the threshold [Imbens and Lemieux, 2008]. Figure 9 and columns 2, 3 and 4 of Table 11 show that there are not discontinuities for the observable demographic characteristics of the children around the threshold.

Moreover, one may be concerned that there is some manipulation of the running variable around the threshold, so that students or their parents coordinate among themselves in order to reach the minimum amount of applications to form a class. In this case, the RDD results may be biased, because it is likely that observations above and below the threshold are different along some unobservable characteristics: parents making applications that fall below the threshold may care less or be less organized, for instance, than parents who manage to coordinate and apply to schools where there are enough TP applications. In this setting, a manipulation of the running variable would require that parents knew the number of students who applied to TP schemes in each school and managed to coordinate applications accordingly. Although this is probably unlikely to happen, we formally test for the presence of manipulation of the running variable using the test designed evaluating whether there is a discontinuity at the cut-off in the density function of the running variable. Figure 10 confirms that the population density is smooth across the cut-off.

8 Conclusions

This paper studies the effect of spending more time at school with respect to independent study on the distribution and dispersion of test scores of students aged 7 and 10. It also investigates whether increasing time spent studying at school - in a homogeneous environment - and reducing time spent studying at home - where the quality of inputs varies greatly and family resources matter - affects achievement gaps by socio-economic background.

It exploits the fact that in Italy two instruction schemes that entail a different amount and organization of time coexist in primary schools: the TN scheme (27 hours per week, usually) and the TP scheme (40 hours per week). Identification relies on arguably supply-driven variation within school-grade in the share of TP classes over time, across subsequent cohorts of second and fifth graders. There are four main findings. First, a larger share of TP classes has a virtually null impact on reading mean test scores, while it positively and significantly affects mathematics mean test scores. Second, while the effect on reading is relatively homogeneous across the distribution, the effect on mathematics is stronger at the bottom than at the top. This implies that the reduction in the inequality of learning outcomes is larger for mathematics. Third, there is evidence of a narrowing of achievement gaps by socio-economic background, although coefficients are not always precisely estimated.

All these effects are stronger in fifth than in second grade. This may suggest the presence of cumulative effects. Moreover, longer school days may increase fatigue and decrease alertness ((OECD [2016a]): older pupils may be more able to adjust to them. The stronger impact on inequality of learning outcomes and achievement gaps may indicate that parental inputs and the broader quality of learning opportunities available outside school grow in importance as pupils age and are exposed to more challenging curricula.

We also shed some light on some mechanisms that may underlie our findings: we show that attending TP schemes reduces the amount of time spent revising the curriculum at home, whereas it does not impact the amount of time devoted to leisure. Overall, total time devoted to instruction varies little, indicating that the substitution between independent and supervised study is close to 1:1. A further line of research could explore how having attended TN or TP schemes affects achievement in secondary school, where TP students may have more difficulties since study gradually becomes more and more autonomous.

To our knowledge, this is the first paper that provides detailed evidence on the causal effects of longer school days in primary school on different percentiles of test scores distribution, on achievement gaps by parental background and on students' time use. We believe that this is of high policy relevance both for the discussion on the way to successfully level the playing filed for disadvantaged pupils - at school and in the society in general - and for the debate on the high share of low performers in reading and, particularly, in mathematics in international standardized tests (OECD [2012]).

References

Jorge M Agüero and Trinidad Beleche. Test-Mex: Estimating the Effects of School Year Length on Student Performance in Mexico. *Journal of Development Economics*, 103: 353–361, 2013.

Joshua D Angrist, Erich Battistin, and Daniela Vuri. In a Small Moment: Class Size and Moral Hazard in the Mezzogiorno. 2014.

Esteban Aucejo and Teresa Romano. Assessing the Effect of School Days and Absences on Test Score Performance. 2014.

Rosario Maria Ballatore, Margherita Fort, and Andrea Ichino. The tower of babel in the classroom: immigrants and natives in italian schools. 2015.

Erich Battistin and Simone Schüller. Evaluation of the "Maestro Unico" Reform. Technical report, IRVAPP, 2013.

Cristián Bellei. Does lengthening the school day increase students' academic achievement? results from a natural experiment in chile. *Economics of Education Review*, 28(5):629–640, 2009.

David Card and Alan B Krueger. Does school quality matter? returns to education and the characteristics of public schools in the united states. *Journal of Political Economy*, 100(1):1–40, 1992.

Eric Eide and Mark H Showalter. The effect of school quality on student performance: A quantile regression approach. *Economics letters*, 58(3):345–350, 1998.

Patrizia Falzetti, Sergio Longobardi, and Paolo Sestito. A Multistep Approach to Detect and Correct the Cheating in Italian Students' Data. Technical report, Meeting of the Italian Statistical Sociey, Statistics and Demography: the Legacy of Corrado Gini, 2015.

Joshua Goodman. Flaking out: Student Absences and Snow Days as Disruptions of Instructional Time. Technical report, National Bureau of Economic Research, 2014.

Jeff Grogger. Does school quality explain the recent black/white wage trend? *Journal of Labor Economics*, pages 231–253, 1996.

Benjamin Hansen. School year length and student performance: Quasi-experimental evidence. *Available at SSRN 2269846*, 2011.

Caroline Hoxby. Peer Effects in the Classroom: Learning from Gender and Race Variation. 2000.

Mathias Huebener, Susanne Kuger, and Jan Marcus. Increased instruction hours and the widening gap in student performance. 2016.

Guido W. Imbens and Thomas Lemieux. Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics*, 142(2):615–635, February 2008.

Victor Lavy. Expanding School Resources and Increasing Time on Task: Effects of a Policy Experiment in Israel on Student Academic Achievement and Behavior. Technical report, National Bureau of Economic Research, 2012.

Victor Lavy. Do Differences in Schools' Instruction Time Explain International Achievement Gaps? Evidence from Developed and Developing Countries. *The Economic Journal*, 125(588):F397–F424, 2015.

Victor Lavy, M Daniele Paserman, and Analia Schlosser. Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom*. *The Economic Journal*, 122(559):208–237, 2012.

Jong-Wha Lee and Robert J Barro. Schooling Quality in a Cross–section of Countries. *Economica*, 68(272):465–488, 2001.

Dave E Marcotte. Schooling and Test Scores: A Mother-natural Experiment. *Economics of Education Review*, 26(5):629–640, 2007.

Dave E Marcotte and Steven W Hemelt. Unscheduled School Closings and Student Performance. *Education*, 3(3):316–338, 2008.

Vincenzo Mariani, Pasquale Recchia, Paolo Sestito, and Valerio Vacca. "Tempo Pieno", Test Scores and Behaviour. Technical report, mimeo, 2012.

OECD. Results in focus: What 15-year-olds know and what they can do with what they know, 2012.

OECD. Student learning time. 2016a.

OECD. How is learning time organised in primary and secondary education? 2016b.

Marco Paccagnella and Paolo Sestito. School cheating and social capital. *Education Economics*, 22(4):367–388, August 2014.

Rasyad A Parinduri. Do Children Spend Too Much Time in Schools? Evidence from a Longer School Year in Indonesia. *Economics of Education Review*, 41:89–104, 2014.

Santiago Pereda Fernández. Teachers and Cheaters. Just an Anagram? Technical report, Mimeo, 2015.

Jörn-Steffen Pischke. The Impact of Length of the School Year on Student Performance and Earnings: Evidence From the German Short School Years. *The Economic Journal*, 117(523):1216–1242, 2007.

Claudio Quintano, Rosalia Castellano, and Sergio Longobardi. A Fuzzy Clustering Approach to Improve the Accuracy of Italian Student Data: An Experimental Procedure to Correct the Impact of Outliers on Assessment Test Scores. 2009.

Ronald Rizzuto and Paul Wachtel. Further evidence on the returns to school quality. *Journal of Human Resources*, pages 240–254, 1980.

Ludger Wößmann. Schooling Resources, Educational Institutions and Student Performance: the International Evidence. *Oxford bulletin of economics and statistics*, 65(2): 117–170, 2003.

Figure 1: Applications for TP classes at the provincial level (scholastic year 2014/2015).

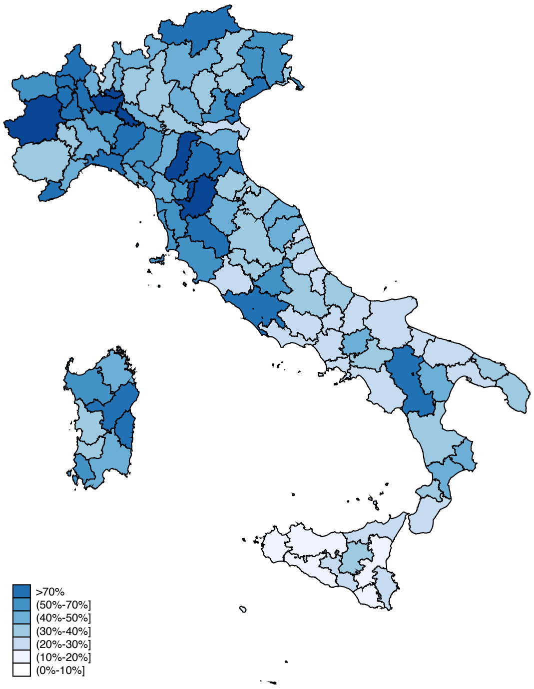

Figure 2: Enrollment in TP classes at the provincial level (scholastic year 2014/2015).

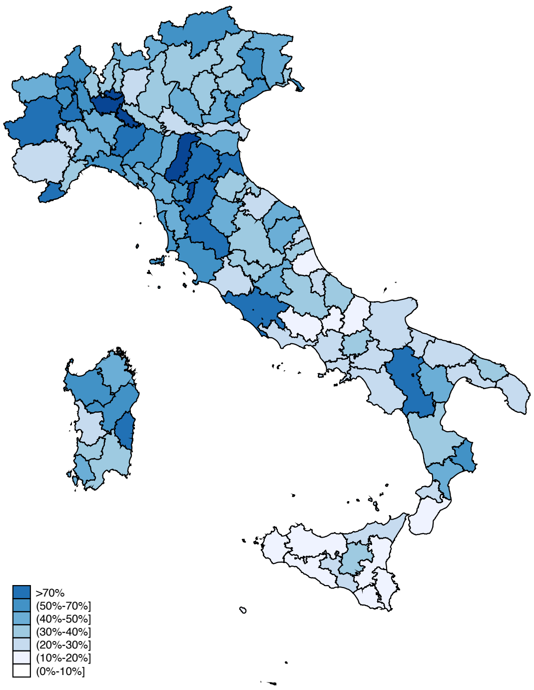

Figure 3: Unmet demand for TP classes at the provincial level (scholastic year 2014/2015).

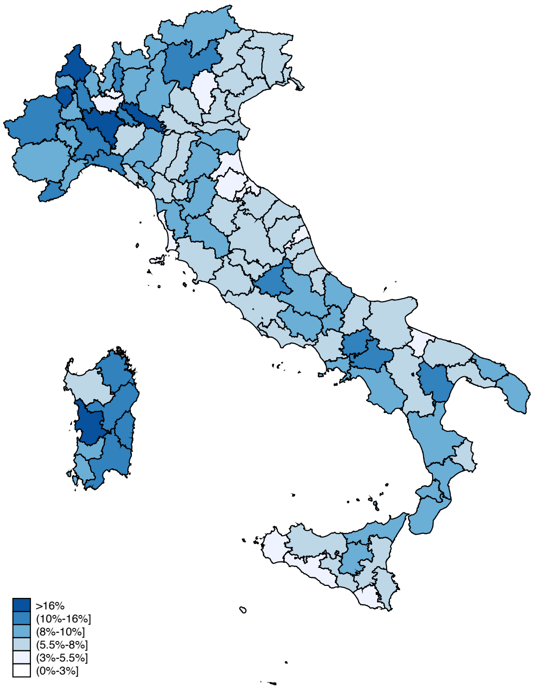

Figure 4: TP and the distributions of reading test scores

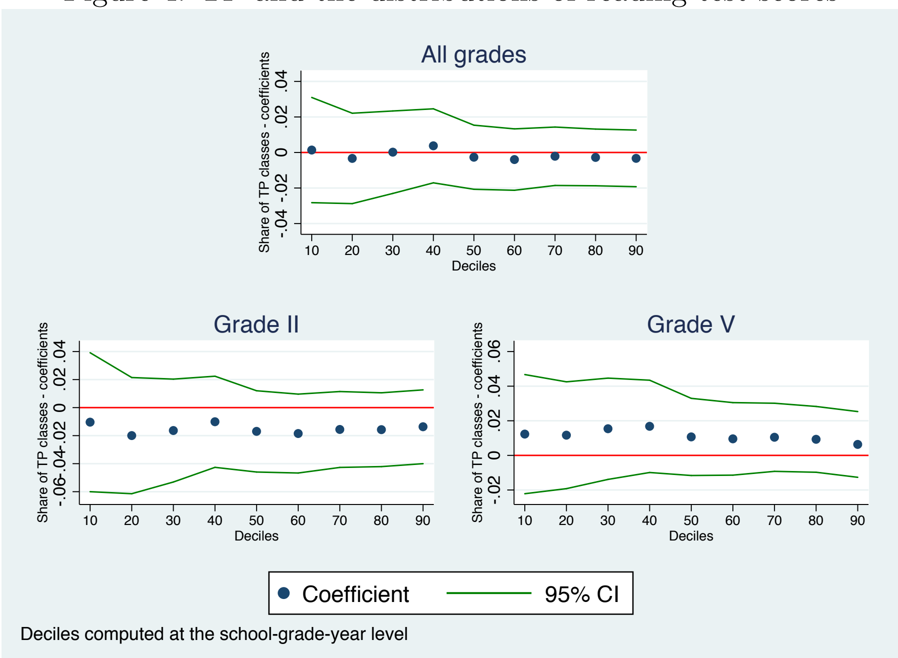

Figure 5: TP and the distribution of mathematics test scores

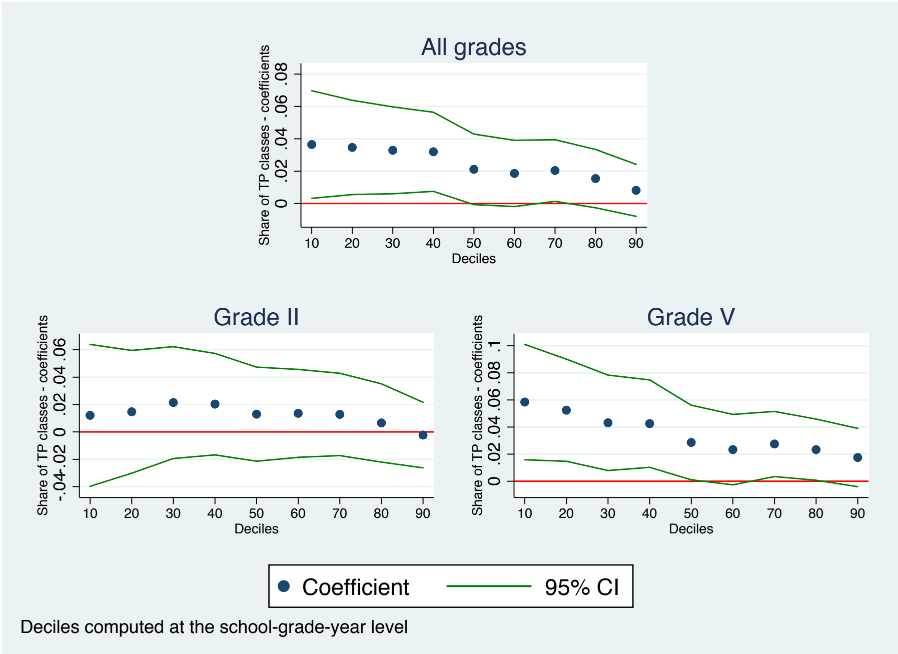

Figure 6: Total time devoted to instruction, in daily minutes

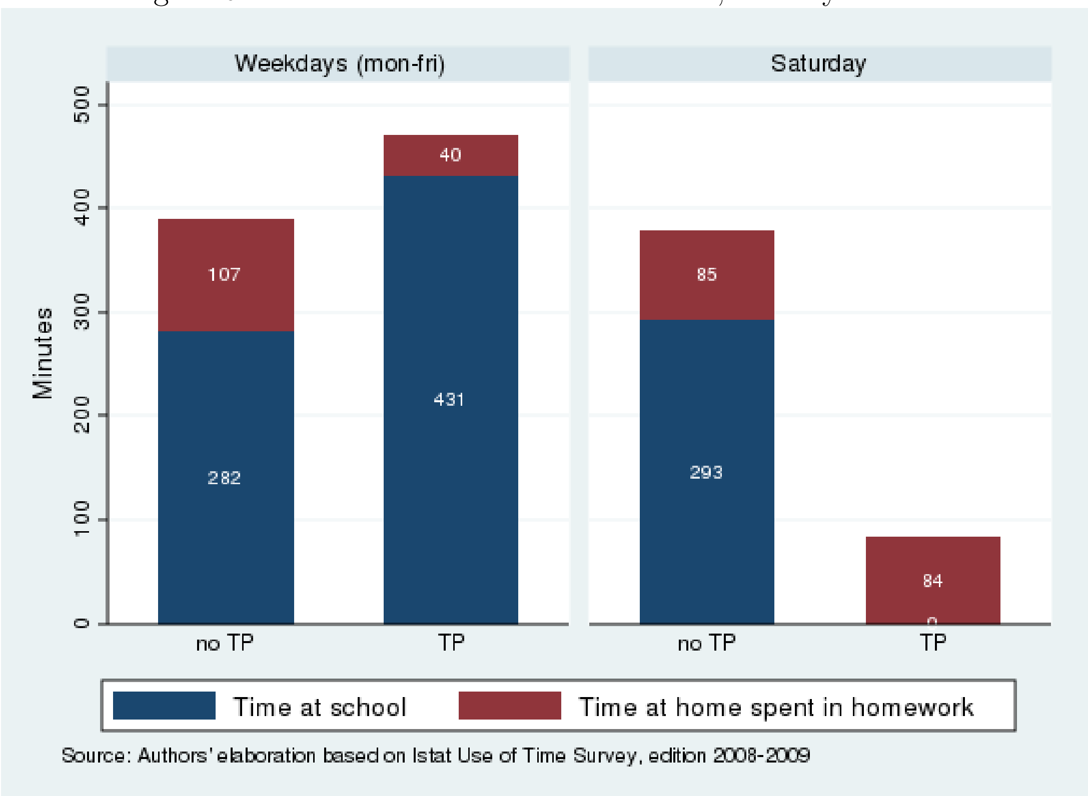

Figure 7: Time spent doing homework, in daily minutes

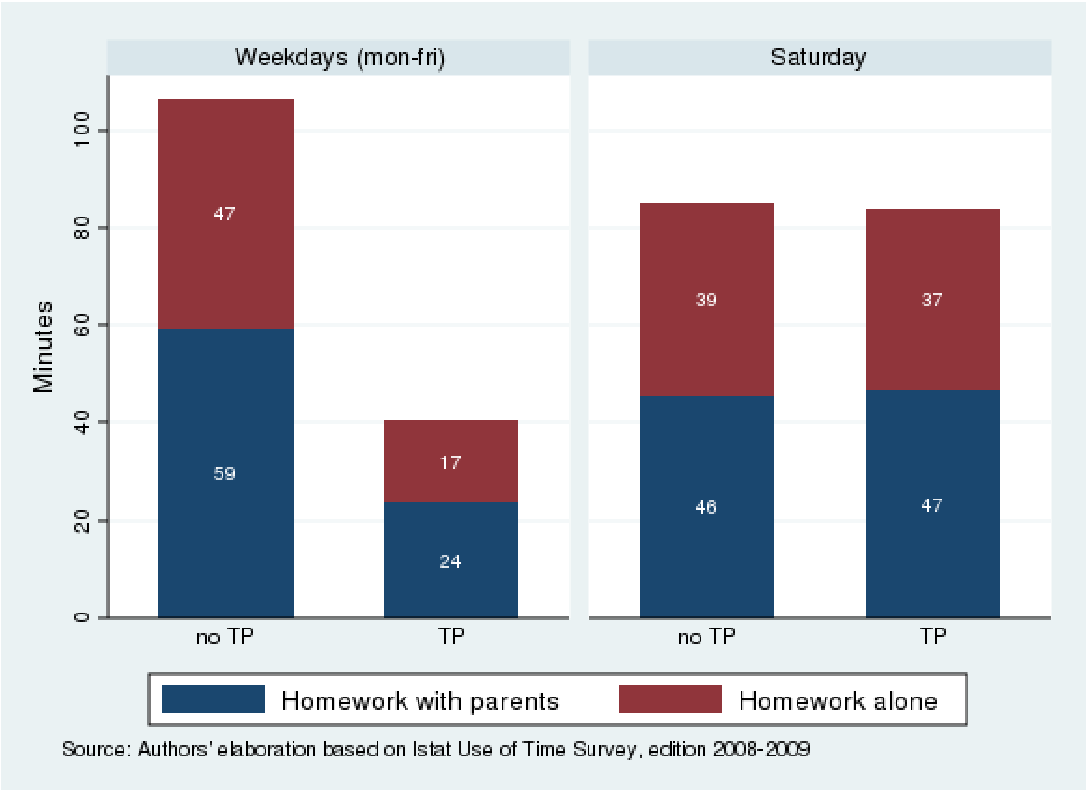

Figure 8: Regression Discontinuity Design - First stages

Comuni non montani

Comuni montani

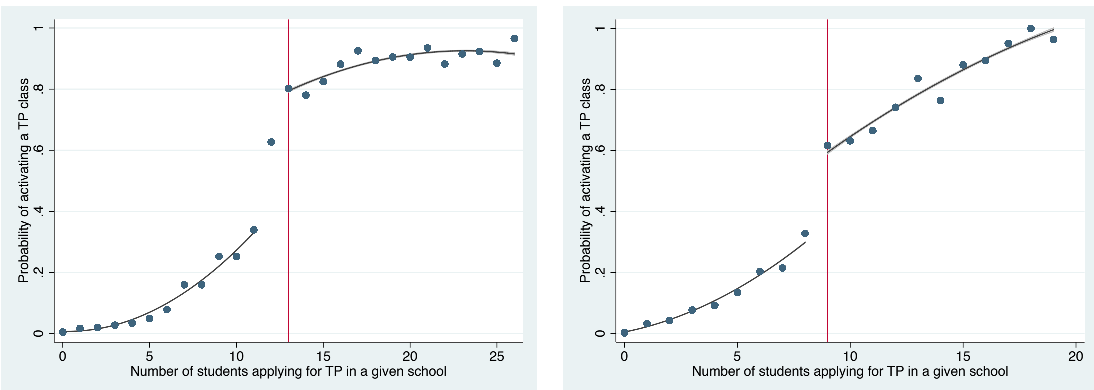

Figure 9: Balancing tests, comuni non montani

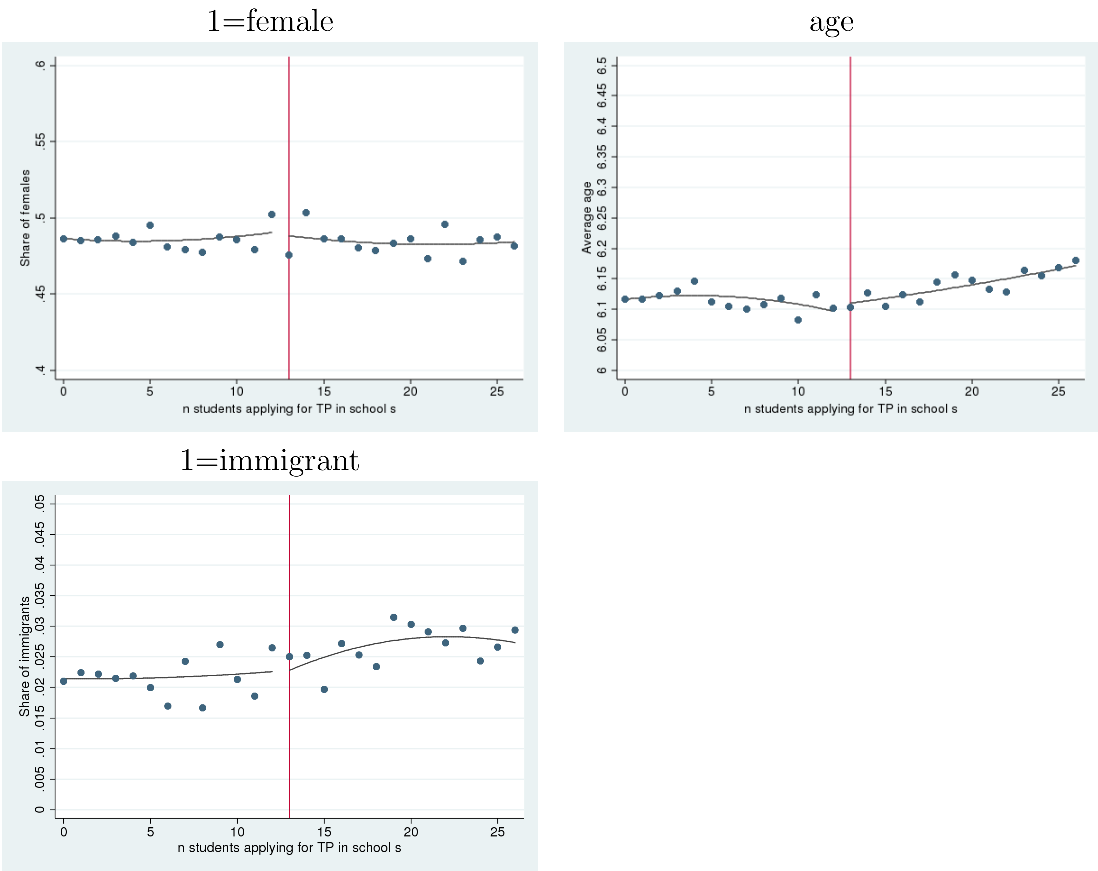

Figure 10: Density checks

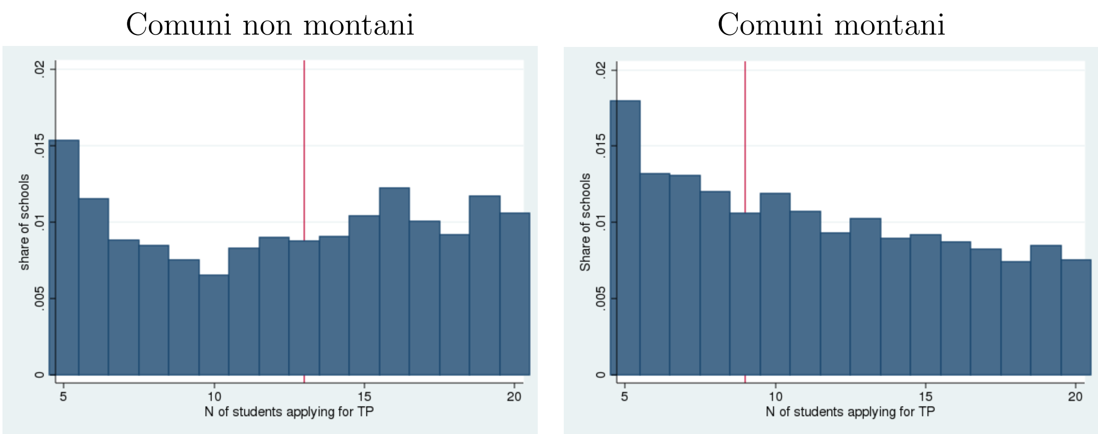

Table 1: Descriptive statistics - Sample

	Grade II		Grade V	
	Mean	SD	Mean	SD
I. School characteristics				
Average enrolment (by grade)	33.76	25.97	33.60	25.83
Average class size	17.81	5.86	17.66	5.78
N. of classes	1.77	1.08	1.78	2.08
Share of TP classes	0.33	0.41	0.29	0.41
II. Students characteristics				
Share male *	0.51	0.50	0.51	0.50
Share native *	0.90	0.31	0.90	0.30
Share regular *	0.97	0.18	0.95	0.21
Share speak prevalently Italian at home *	.	.	0.92	0.27
Share attended nursery *	0.35	0.48	0.30	0.46
Share father with HS diploma or more *	0.51	0.50	0.48	0.50
Share mother with HS diploma or more *	0.60	0.49	0.56	0.50
Share mother not working *	0.45	0.50	0.46	0.50
Average ESCS *	.	.	0.04	1.03

Note: The unit of observation in Panel I is the school-year. The unit of observation in Panel II is the student. The ESCS indicator summarizes information about parents' education, occupation and wealth. It is standardized by grade and year to have zero mean and unit standard deviation. Information about the language prevalently spoken at home and the ESCS indicator is available for fifth graders only. * Shares are conditional on non-missing survey response.

Table 2: TN vs TP: students and parents characateristics

	TN	TP	Diff: TN - TP
I. Local area characteristics			
Average population	142841	375855	-233014.42
Share in provincial capital	0.25	0.37	-0.11
Average unemp. rate (2011-2014)	0.13	0.10	0.03
Average female non part. rate (2011-2013)	0.29	0.20	0.09
II. School and student characteristics			
Average class size	17.27	19.00	-1.73
Share male *	0.51	0.51	-0.00
Share with Italian citizenship *	0.92	0.85	0.06
Share regular *	0.96	0.97	-0.01
Share speak prevalently Italian/Dialect at home *	0.93	0.90	0.04
Share attended nursery *	0.28	0.42	-0.14
Share father with HS diploma or more *	0.49	0.53	-0.04
Share mother with HS diploma or more *	0.56	0.62	-0.05
Share mother not working *	0.50	0.34	0.16
Average ESCS *	0.02	0.11	-0.09
II. Test scores			
Average reading test score (NC)	0.02	-0.04	0.06
Average reading test score (C)	-0.02	0.03	-0.05
Average mathematics test score (NC)	0.02	-0.04	0.07
Average mathematics test score (C)	-0.01	0.02	-0.03

Note: The ESCS indicator summarizes information about parents' education, occupation and wealth. It is standardized by grade and year to have zero mean and unit standard deviation. Information about the language prevalently spoken at home and the ESCS indicator comes from the questionnaire administered to fifth graders only. Test scores are standardized by subject, grade and year to have zero mean and unit standard deviation. Scores not corrected for cheating (NC) are available from 2010-2011 to 2014-2015. Scores corrected for cheating (C) are available from 2011-2012 to 2014-2015. * Shares are conditional on non-missing survey response.

Table 3: Balancing tests

	[1]	[2]	[3]	[4]	[5]	[6]
	Enrollment			# absent		
Share classes TP	5.905***	-0.728***	-1.000***	1.038***	0.369***	0.071
	(0.315)	(0.226)	(0.289)	(0.042)	(0.075)	(0.094)
Obs.	118662	118662	118662	118662	118662	118662
Y mean		39.86			3.25	
	Share male			Share gender missing		
Share classes TP	0.002***	0.004	0.002	0.000	0.000	0.000
	(0.001)	(0.002)	(0.004)	(0.000)	(0.000)	(0.000)
Obs.	118662	118662	118662	118662	118662	118662
Y mean		0.51			0.00	
	Share with Italian citizenship			Share citizenship missing		
Share classes TP	-0.054***	-0.006***	-0.003	0.001***	-0.000	0.000
	(0.002)	(0.002)	(0.003)	(0.000)	(0.000)	(0.001)
Obs.	118633	118633	118633	118662	118662	118662
Y mean		0.90			0.00	
	Share regular			Share regularity missing		
Share classes TP	0.005***	0.003***	-0.002	0.000***	0.002**	0.003**
	(0.001)	(0.001)	(0.002)	(0.000)	(0.001)	(0.001)
Obs.	118641	118641	118641	118662	118662	118662
Y mean		0.96			0.00	
	Share mother with HS dip. at least			Share edu. mother missing		
Share classes TP	0.043***	0.018***	-0.002	0.040***	-0.008	-0.013
	(0.002)	(0.003)	(0.004)	(0.003)	(0.005)	(0.008)
Obs.	106403	106403	106403	118662	118662	118662
Y mean		0.58			0.18	
	Share father with HS dip. at least			Share edu. father missing		
Share classes TP	0.037***	0.014***	-0.002	0.043***	-0.008	-0.012
	(0.003)	(0.003)	(0.005)	(0.003)	(0.005)	(0.008)
Obs.	106149	106149	106149	118662	118662	118662
Y mean		0.49			0.20	
School × *grade* FE	No	Yes	Yes	No	Yes	Yes
School × *grade* trends	No	No	Yes	No	No	Yes

Note: The unit of observation is the school-grade-year. The independent variable is the share of TP classes. The dependent variables are, from the left to the right and from the top to the bottom: enrollment, the number of students absent during at least one test, the share of male students, the share of students for whom gender is missing, the share of students with Italian citizenship, the share of students for whom citizenship is missing, the share of regular (i.e. not ahead or behind the same-cohort grade level) students, the share of students for whom regularity is missing, the share of mothers and fathers with at least the high-school diploma, the share of mothers and fathers for whom education is missing. All regressions include year fixed effects. Columns 1 and 4 report estimates without including school-grade fixed effects. Columns 2 and 5 report estimates augmented with school-grade fixed effects. Columns 3 and 6 report estimates augmented with school-grade fixed effects and school-grade specific linear time trends. Standard errors are clustered at the school-grade level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: TP and the distribution of Reading test scores

	d10	d20	d30	d40	d50	d60	d70	d80	d90
					All grades				
Share classes TP	0.001	-0.003	0.000	0.004	-0.003	-0.004	-0.002	-0.003	-0.003
	(0.015)	(0.013)	(0.012)	(0.011)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)
Obs.	81637	81637	81637	81637	81637	81637	81637	81637	81637
					Grade II				
Share classes TP	-0.010	-0.020	-0.016	-0.010	-0.017	-0.019	-0.016	-0.016	-0.014
	(0.025)	(0.021)	(0.019)	(0.017)	(0.015)	(0.014)	(0.014)	(0.013)	(0.013)
Obs.	38403	38403	38403	38403	38403	38403	38403	38403	38403
					Grade V				
Share classes TP	0.012	0.012	0.015	0.017	0.011	0.010	0.010	0.009	0.006
	(0.018)	(0.016)	(0.015)	(0.014)	(0.011)	(0.011)	(0.010)	(0.010)	(0.010)
Obs.	43234	43234	43234	43234	43234	43234	43234	43234	43234
Student controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School × *grade* FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School × *grade* trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The unit of observation is the school-grade-year. The independent variable of interest is the share of TP classes. The dependent variables are the (log) deciles of reading test scores (corrected for cheating). The set of controls include: the share of male students, the share of students with Italian citizenship, the share of regular students, the share of mothers and fathers with at least the high-school diploma, along with the share of students who report missing values in each of these variables. They also include: the enrolment and its square, the number of classes, the share of students absent the day of at least one test, a dummy for whether an external invigilator proctored the test in at least one class, a dummy for whether the Gelmini reform was in place. All regressions also include year fixed effects, school-grade fixed effects and school-grade linear trends. Standard errors are clustered at the school-grade level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: TP and the distribution of Mathematics test scores

	d10	d20	d30	d40	d50	d60	d70	d80	d90
					All grades				
Share classes TP	0.036**	0.035**	0.033**	0.032**	0.021*	0.019*	0.020**	0.015*	0.008
	(0.017)	(0.015)	(0.014)	(0.012)	(0.011)	(0.010)	(0.010)	(0.009)	(0.008)
Obs.	81682	81682	81682	81682	81682	81682	81682	81682	81682
					Grade II				
Share classes TP	0.012	0.015	0.021	0.020	0.013	0.014	0.013	0.007	-0.002
	(0.026)	(0.023)	(0.021)	(0.019)	(0.018)	(0.016)	(0.015)	(0.015)	(0.012)
Obs.	38434	38434	38434	38434	38434	38434	38434	38434	38434
					Grade V				
Share classes TP	0.058***	0.052***	0.043**	0.043***	0.029**	0.023*	0.027**	0.023**	0.017
	(0.022)	(0.019)	(0.018)	(0.016)	(0.014)	(0.013)	(0.012)	(0.012)	(0.011)
Obs.	43248	43248	43248	43248	43248	43248	43248	43248	43248
Student controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School × *grade* FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School × *grade* trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The unit of observation is the school-grade-year. The independent variable of interest is the share of TP classes. The dependent variables are the (log) deciles of mathematics test scores (corrected for cheating). Time-varying controls are the same as in Table 4. All regressions also include year fixed effects, school-grade fixed effects and school-grade linear trends. Standard errors are clustered at the school-grade level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Measures of central tendency

	Reading			Mathematics		
	All grades	Grade II	Grade V	All grades	Grade II	Grade V
				Mean score		
Share classes TP	-0.002	-0.013	0.009	0.022**	0.011	0.032**
	(0.009)	(0.014)	(0.011)	(0.010)	(0.015)	(0.013)
				Median score		
Share classes TP	-0.004	-0.016	0.008	0.021*	0.013	0.029**
	(0.009)	(0.015)	(0.012)	(0.011)	(0.018)	(0.014)
Student controls	Yes	Yes	Yes	Yes	Yes	Yes
School controls	Yes	Yes	Yes	Yes	Yes	Yes
School × *grade* FE	Yes	Yes	Yes	Yes	Yes	Yes
School × *grade* trends	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	81696	38442	43254	81698	38442	43256

Note: The unit of observation is the school-grade-year. The independent variable of interest is the share of TP classes. The dependent variables are the (log) mean and the median test scores (corrected for cheating). Time-varying controls are the same as in Table 4. All regressions also include year fixed effects, school-grade fixed effects and school-grade linear trends. Standard errors are clustered at the school-grade level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Measures of dispersion

	Reading			Mathematics		
	All grades	Grade II	Grade V	All grades	Grade II	Grade V
			Std. Dev.			
Share classes TP	-0.019	-0.013	-0.023	-0.015	-0.004	-0.025
	(0.018)	(0.026)	(0.024)	(0.015)	(0.020)	(0.021)
			Ratio 75/25			
Share classes TP	0.001	0.007	-0.005	-0.016	-0.008	-0.023*
	(0.009)	(0.014)	(0.012)	(0.010)	(0.015)	(0.013)
			Ratio 90/10			
Share classes TP	-0.003	-0.001	-0.006	-0.028*	-0.014	-0.041**
	(0.014)	(0.024)	(0.015)	(0.014)	(0.023)	(0.019)
Student controls	Yes	Yes	Yes	Yes	Yes	Yes
School controls	Yes	Yes	Yes	Yes	Yes	Yes
School × *grade* FE	Yes	Yes	Yes	Yes	Yes	Yes
School × *grade* trends	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	81545	38353	43192	81611	38396	43215

Note: The unit of observation is the school-grade-year. The independent variable of interest is the share of TP classes. The dependent variables are the (log) standard deviation, the 75/25 ratio and the 90/10 ratio of test scores (corrected for cheating). Time-varying controls are the same as in Table 4. All regressions also include year fixed effects, school-grade fixed effects and school-grade linear trends. Standard errors are clustered at the school-grade level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Achievement gaps

	Reading				Mathematics			
	Mother education	Father education	Citizenship	Language at home	Mother education	Father education	Citizenship	Language at home
				All grades				
Share classes TP	-0.004	-0.008	-0.003	-0.014	-0.010	-0.006	-0.006	-0.020
	(0.007)	(0.007)	(0.014)	(0.020)	(0.009)	(0.008)	(0.015)	(0.023)
Obs.	76836	77144	49782	25066	76928	77198	50004	25471
				Grade II				
Share classes TP	0.001	0.001	-0.001		-0.004	0.009	0.006	
	(0.012)	(0.011)	(0.023)		(0.014)	(0.012)	(0.024)	
Obs.	36131	36384	24039		36160	36408	24145	
				Grade V				
Share classes TP	-0.008	-0.016*	-0.002	-0.014	-0.015	-0.019*	-0.017	-0.020
	(0.008)	(0.008)	(0.017)	(0.020)	(0.012)	(0.010)	(0.020)	(0.023)
Obs.	40705	40760	25743	25066	40768	40790	25859	25471
Student controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School × *grade* FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School × *grade* trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The unit of observation is the school-grade-year. The independent variable of interest is the share of TP classes. The dependent variables are the (log) achievement gaps, expressed as ratios, by mother education (diploma/not diploma), father education (diploma/not diploma), student citizenship (Italian/foreign) and language prevalently spoken at home (Italian or regional dialect/foreign language). The information about the language prevalently spoken at home is available for fifth graders only. Time-varying controls are the same as in Table 4. All regressions also include year fixed effects, school-grade fixed effects and school-grade linear trends. Standard errors are clustered at the school-grade level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: TN vs TP: use of time at home

	TN	TP	Diff: TN - TP
Share never do homework *	0.02	0.03	-0.01
Share do homework 1-2 times a week *	0.12	0.37	-0.25
Share do homework 3-4 times a week *	0.22	0.36	-0.13
Share do homework more than 5 times a week *	0.64	0.24	0.40
Share watch TV more than 1 hr a day *	0.52	0.53	-0.01
Share play with PC/videogames more than 1 hr a day *	0.47	0.46	0.01
Share play with friends more than 1 hr a day *	0.81	0.81	-0.00
Share help with housework more than 1 hr a day *	0.42	0.40	0.02
Share read books/comics more than 1 hr a day *	0.31	0.31	0.01
Share play sport more than 3 times a week *	0.38	0.36	0.02
Share do other recr. act. more than 3 times a week *	0.08	0.06	0.02

Note: The unit of observation is the student. Information about time use is collected up to the 2012-2013 wave of the student questionnaire. * Shares are conditional on non-missing survey responses.

Table 10: Use of time at home

	Coeff.	Obs.
Never homework	0.011***	35616
	(0.004)	
Homework $<$ 2 times/week	0.058***	35616
	(0.016)	
Homework $>$ 5 times/week	-0.073***	35616
	(0.017)	
Watch TV $>$ 1 hr/day	-0.017	35618
	(0.014)	
Play w/PC $>$ 1 hr/day	-0.020	35617
	(0.014)	
Play w/ friends $>$ 1 hr/day	-0.009	35617
	(0.010)	
Housework $>$ 1 hr/day	0.011	35617
	(0.012)	
Read books $>$ 1 hr/day	0.015	35617
	(0.012)	
Play sport $>$ 3 times/week	0.000	35617
	(0.012)	
Other recreational activities $>$ 3 times/week	0.002	35615
	(0.007)	
Student controls	Yes	
School controls	Yes	
School $\times$ *grade* FE	Yes	
School $\times$ *grade* trends	Yes	

Note: The unit of observation is the school-grade-year. The independent variable of interest is the share of TP classes. The dependent variables are the share of students who: never do homework; do homework less than twice a week;do homework more than 5 times a week; watch TV more than 1 hour per day; play with PC more than 1 hour per day; play with friends more than 1 hour per day; do housework more than 1 hour per day; read books more than 1 hour per day; play sport or engage in other recreational activities more than 3 times per week. Time-varying controls are the same as in Table 4. All regressions also include year fixed effects, school-grade fixed effects and school-grade linear trends. Standard errors are clustered at the school-grade level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 11: Regression Discontinuity Design: first stage and balancing tests

	Activate TP class	1= female	1= immigrant	Age
	Entire sample, 3 degree polynomial			
TP applicants>min size	0.135***	0.003	-0.002	-0.012
	(0.016)	(0.004)	(0.002)	(0.010)
TP applicants	0.024***	-0.000	0.000***	0.002***
	(0.001)	(0.000)	(0.000)	(0.001)
TP applicants^2	-0.000***	0.000	-0.000*	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
TP applicants^3	0.000***	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Obs.	525280	525280	525280	525280
	Around the threshold, 2 degree polynomial			
TP applicants>min size	0.091***	0.004	-0.001	-0.014
	(0.018)	(0.005)	(0.002)	(0.014)
TP applicants	0.025***	-0.000	0.000	0.000
	(0.001)	(0.001)	(0.000)	(0.002)
TP applicants^2	-0.000**	0.000	0.000*	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Obs.	346636	346636	346636	346636

Note: The unit of observation is the student. The independent variable of interest is whether the number of applicants in school s is above the minimum class size threshold. The dependent variables are: a dummy for whether the school offers at least on TP class, a female dummy, an immigrant dummy and the age (in months). Time-varying controls are the same as in Table 4. Column 1 also includes demographic controls. Standard errors are clustered at the school level. The second panel includes only schools for wichthe number of TP applicants is below 25. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

A Appendix

Figure 11: TP and the distributions of reading test scores (not corrected for cheating)

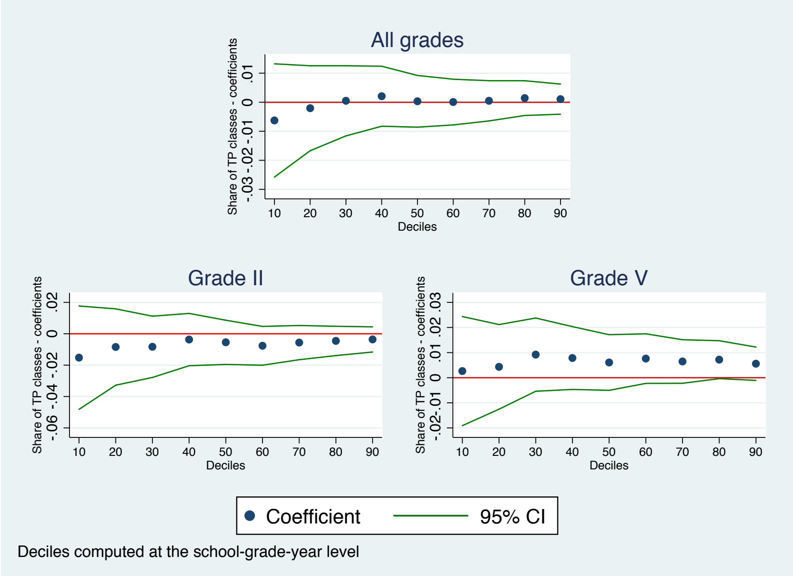

Figure 12: TP and the distribution of mathematics test scores (not corrected for cheating)

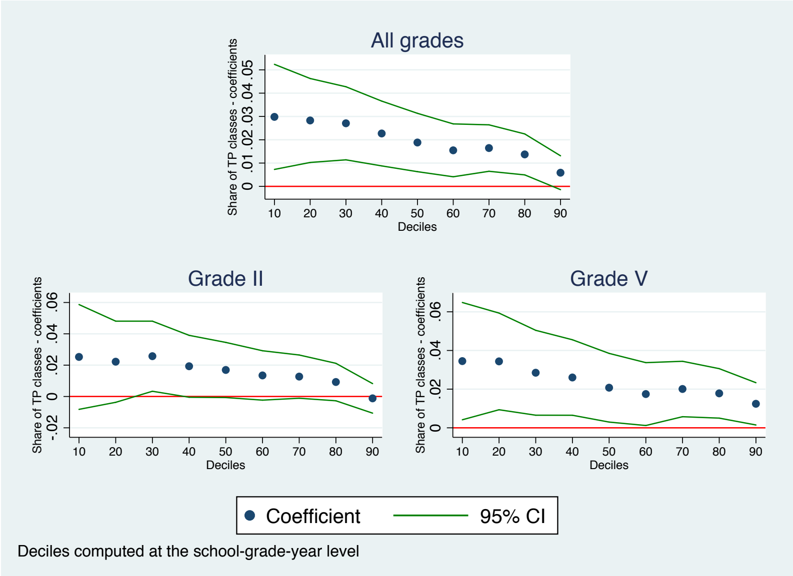

Table 12: TP and the distribution of reading test scores (scores not corrected for cheating)

	d10	d20	d30	d40	d50	d60	d70	d80	d90
					All grades				
Share classes TP	-0.006	-0.002	0.000	0.002	0.000	0.000	0.000	0.001	0.001
	(0.010)	(0.007)	(0.006)	(0.005)	(0.005)	(0.004)	(0.004)	(0.003)	(0.003)
Obs.	105688	105688	105688	105688	105688	105688	105688	105688	105688
					Grade II				
Share classes TP	-0.015	-0.008	-0.008	-0.004	-0.005	-0.008	-0.006	-0.005	-0.004
	(0.017)	(0.012)	(0.010)	(0.009)	(0.007)	(0.006)	(0.006)	(0.005)	(0.004)
Obs.	49114	49114	49114	49114	49114	49114	49114	49114	49114
					Grade V				
Share classes TP	0.003	0.004	0.009	0.008	0.006	0.008	0.006	0.007*	0.006
	(0.011)	(0.009)	(0.007)	(0.006)	(0.006)	(0.005)	(0.004)	(0.004)	(0.003)
Obs.	56574	56574	56574	56574	56574	56574	56574	56574	56574
Student controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School × *grade* FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School × *grade* trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The unit of observation is the school-grade-year. The independent variable of interest is the share of TP classes. The dependent variables are the (log) deciles of the distribution of reading test scores (not corrected for cheating). Time-varying controls are the same as in Table 4. All regressions also include year fixed effects, school-grade fixed effects and school-grade linear trends. Standard errors are clustered at the school-grade level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 13: TP and the distribution of mathematics test scores (scores not corrected for cheating)

	d10	d20	d30	d40	d50	d60	d70	d80	d90
					All grades				
Share classes TP	0.030***	0.028***	0.027***	0.023***	0.019***	0.015***	0.016***	0.014***	0.006
	(0.012)	(0.009)	(0.008)	(0.007)	(0.006)	(0.006)	(0.005)	(0.004)	(0.004)
Obs.	105722	105722	105722	105722	105722	105722	105722	105722	105722
					Grade II				
Share classes TP	0.025	0.022*	0.026**	0.019*	0.017*	0.013*	0.013*	0.009	-0.001
	(0.017)	(0.013)	(0.011)	(0.010)	(0.009)	(0.008)	(0.007)	(0.006)	(0.005)
Obs.	49139	49139	49139	49139	49139	49139	49139	49139	49139
					Grade V				
Share classes TP	0.034**	0.034***	0.028**	0.026***	0.021**	0.017**	0.020***	0.018***	0.012**
	(0.015)	(0.013)	(0.011)	(0.010)	(0.009)	(0.008)	(0.007)	(0.007)	(0.006)
Obs.	56583	56583	56583	56583	56583	56583	56583	56583	56583
Student controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School × *grade* FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School × *grade* trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The unit of observation is the school-grade-year. The independent variable of interest is the share of TP classes. The dependent variables are the (log) deciles of the distribution of mathematics test scores (not corrected for cheating). Time-varying controls are the same as in Table 4. All regressions also include year fixed effects, school-grade fixed effects and school-grade linear trends. Standard errors are clustered at the school-grade level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 14: Measures of central tendency (scores not corrected for cheating)

	Reading			Mathematics		
	All grades	Grade II	Grade V	All grades	Grade II	Grade V
			Mean score			
Share classes TP	-0.000	-0.006	0.005	0.018***	0.013*	0.022***
	(0.004)	(0.007)	(0.005)	(0.006)	(0.008)	(0.008)
			Median score			
Share classes TP	-0.001	-0.006	0.005	0.019***	0.017*	0.021**
	(0.005)	(0.007)	(0.006)	(0.006)	(0.009)	(0.009)
Student controls	Yes	Yes	Yes	Yes	Yes	Yes
School controls	Yes	Yes	Yes	Yes	Yes	Yes
School $\times$ *grade* FE	Yes	Yes	Yes	Yes	Yes	Yes
School $\times$ *grade* trends	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	105772	49154	56618	105764	49154	56610

Note: The unit of observation is the school-grade-year. The independent variable of interest is the share of TP classes. The dependent variables are the (log) mean and median test scores (not corrected for cheating). Time-varying controls are the same as in Table 4. All regressions also include year fixed effects, school-grade fixed effects and school-grade linear trends. Standard errors are clustered at the school-grade level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 15: Measures of dispersion (scores not corrected for cheating)

	Reading			Mathematics		
	All grades	Grade II	Grade V	All grades	Grade II	Grade V
			Std. Dev.			
Share classes TP	0.002	0.003	-0.000	-0.007	-0.011	-0.003
	(0.010)	(0.015)	(0.015)	(0.009)	(0.013)	(0.013)
Obs.	105595	49068	56527	105634	49091	56543
			Ratio 75/25			
Share classes TP	0.002	0.006	-0.002	-0.014**	-0.013	-0.015*
	(0.005)	(0.008)	(0.006)	(0.006)	(0.009)	(0.008)
			Ratio 90/10			
Share classes TP	0.008	0.012	0.003	-0.024**	-0.027*	-0.023*
	(0.009)	(0.016)	(0.010)	(0.010)	(0.015)	(0.013)
Student controls	Yes	Yes	Yes	Yes	Yes	Yes
School controls	Yes	Yes	Yes	Yes	Yes	Yes
School × *grade* FE	Yes	Yes	Yes	Yes	Yes	Yes
School × *grade* trends	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	105595	49068	56527	105634	49091	56543

Note: The unit of observation is the school-grade-year. The independent variable of interest is the share of TP classes. The dependent variables are the (log) standard deviation, the (log) 75/25 ratio and the (log) 90/10 ratio of test scores (not corrected for cheating). Time-varying controls are the same as in Table 4. All regressions also include year fixed effects, school-grade fixed effects and school-grade linear trends. Standard errors are clustered at the school-grade level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 16: Achievement gaps (scores not corrected for cheating)

	Reading				Mathematics			
	Mother education	Father education	Citizenship	Language at home	Mother education	Father education	Citizenship	Language at home
				All grades				
Share classes TP	0.004	-0.004	-0.005	-0.014	-0.005	-0.007	-0.004	-0.021
	(0.005)	(0.005)	(0.010)	(0.015)	(0.006)	(0.006)	(0.012)	(0.018)
Obs.	100192	100454	65669	33227	100256	100493	65870	33697
				Grade II				
Share classes TP	0.014*	-0.003	-0.005		-0.002	-0.006	0.003	
	(0.008)	(0.008)	(0.016)		(0.009)	(0.009)	(0.018)	
Obs.	46554	46808	31319		46576	46828	31416	
				Grade V				
Share classes TP	-0.006	-0.004	-0.004	-0.014	-0.008	-0.008	-0.010	-0.021
	(0.006)	(0.006)	(0.012)	(0.015)	(0.009)	(0.008)	(0.016)	(0.018)
Obs.	53638	53646	34350	33227	53680	53665	34454	33697
Student controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School × *grade* FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School × *grade* trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The unit of observation is the school-grade-year. The independent variable of interest is the share of TP classes. The dependent variables are the (log) achievement gaps, expressed as ratios, by mother education (diploma/not diploma), father education (diploma/not diploma), student citizenship (Italian/foreign) and language prevalently spoken at home (Italian or regional dialect/foreign language). Scores are not corrected for cheating. The information about the language prevalently spoken at home is available for fifth graders only. Time-varying controls are the same as in Table 4. All regressions also include year fixed effects, school-grade fixed effects and school-grade linear trends. Standard errors are clustered at the school-grade level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.