Should We Increase Instruction Time in Low Achieving Schools? Evidence from Southern Italy

Erich Battistin*
Queen Mary University of London, CEPR, IRVAPP and IZA

Elena Claudia Meroni†
European Commission, Joint Research Centre

Abstract

This paper investigates the short term effects of a large scale intervention, funded by the European Social Fund, which provides additional instruction time to selected classes of lower secondary schools in Southern Italy. Selection is addressed using institutional rules that regulate class formation: first year students are divided into groups distinguished by letters, they remain in the same group across grades at the school, and the composition of teachers assigned to groups is stable over time. Using a difference-in-differences strategy, we consider consecutive cohorts of first year students enrolled in the same group. We compare participating groups to non-participating groups within the same school, as well as to groups in non-participating schools. We find that the intervention raised scores in mathematics for students from the least advantaged backgrounds. We also find that targeting the best students with extra activities in language comes at the cost of lowering performance in mathematics. We go beyond average effects, finding that the positive effect for mathematics is driven by larger effects for the best students.

JEL Classification: C31,I21, I28

Keywords: Educational economics, instruction time, policy evaluation, quantile treatment effects

*Erich Battistin, School of Economics and Finance Queen Mary, University of London, Mile End Road, London E1 4NS; mail: e.battistin@qmul.ac.uk and

†Corresponding author. Elena Claudia Meroni, European Commission, Joint Research Centre, Via Fermi, 21027 Ispra, Italy; mail: elena.meroni@jrc.ec.europa.eu; phone: +39 0332 785608. The views expressed in this paper are those of the authors and should not be attributed to the European Commission.

1 Introduction

Understanding the key drivers of quality in education plays a fundamental role to meet the EU "Education and Training 2020" targets. The relevance of this problem for policy making is particularly important in areas facing marked socio-economic deprivation. Given the conspicuous investments made to finance structural assistance, providing evidence on the key dimensions that should be targeted by public interventions in Europe adds to the discussion on the most effective growth strategies for the coming decades.

This paper focuses on education policies that mandate low achieving students to extra hours at school. We exploit variation in instruction time resulting from an intervention implemented in selected regions of Southern Italy eligible for EU Regional Development Funds (Objective 1 regions) and EU Social Funds. The Quality and Merit Project (PQM in what follows) started in 2010 and targets low achieving lower secondary schools of these regions (from sixth to eighth grade).[1] The rationale for intervening stems from the fact that schools of these areas are characterised by markedly lower student performance if compared to schools in the rest of the country.

Participation of schools is not mandatory. Applicant schools are ranked using performance indicators (i.e. percentage of repeating and failing students and dropout rates), and only those at the bottom end of the performance distribution are eventually enrolled. Schools admitted to the programme must organise education activities outside regular school hours in a selected number of classes chosen by school principals at the time of application. Teachers involved in afternoon activities must be tenured and with a full-time contract, and teach the same students during normal school hours. Activities vary across students, ranging from remedial classes to study programmes aimed at mastering advanced skills. All students in the class participate in at least one of these activities. Extra time is organised as individualised instruction in small study groups, and all costs are covered through EU funds.

The intervention we consider shifts public spending for the classes involved above the OECD average, and is worth a 7 percent increase of current per-student figures in Italy (9, 100USD, as opposed to 9, 300USD in OECD countries). We use longitudinal information on test scores for two consecutive cohorts of students in PQM schools before and after 2010, and in similar schools located in Objective 1 areas but *not* enrolled in PQM. We use standardised scores for sixth graders at the national level, which we link to administrative information on participating schools to investigate the (short term) effects on learning. Given the lack of standardised scores at the national level before the school year 2009/10, our paper is the first to provide a rigorous evaluation of the effect of the EU

[1]The project is funded through *PON Istruzione 2007-2013* (A-2-FSE-2009-2). The four targeted regions are Apulia, Calabria, Campania and Sicily.

funds spent on education in Italy on a large scale.

We address selection of participating classes using features of the Italian school system. First year students are divided into groups, called *sezione*, and remain in the same group for the whole cycle of studies.[2] Assignment of students to groups is not random. The key feature that we exploit for identification is that teachers are typically assigned to the same *sezione* over school years and across grades. The adoption of this practice is prompted by a reform that was implemented nationwide in 2009 (before the intervention we consider) to regulate staff ratios. For example, the law states that math teachers must fulfil their weekly duties (18 hours in total) by teaching modules of 6 hours to three different classes, and explicitly suggests that these should be the sixth, seventh and eighth grades of the same *sezione* to ensure continuity of teaching practices. We consider a difference-in-differences strategy, and compare changes in test scores for consecutive cohorts of students enrolled in the same *sezione* before and after PQM. As teachers employed in a *sezione* are the same across cohorts, our approach controls for teacher unobservables which are likely drivers of selection into the intervention. We check the sensitivity of our results to the presence of pre-intervention trends in test scores using data for fifth graders. Despite some data limitations that are discussed below, we are not able to reject the validity of our difference-in-differences strategy.

We exploit within school variability in class participation to study effects on students not directly involved in PQM. We are particularly concerned with possible interactions between teachers of participating classes and other teachers at the school, as the former must attend a preparatory course before organising afternoon activities. We assess the existence of indirect effects by comparing students in non-participating classes of PQM schools to students in non-participating schools located in Objective 1 area. In addition, we study heterogeneity in the effects of instruction time along two different dimensions. We let returns depend on number of school hours on top of normal instruction time, thus studying a dose-response model rather than the coefficient on a participation dummy. Moreover, we go beyond average effects and investigate the returns to participation on quantiles of the score distribution. This choice is motivated by the sizeable heterogeneity in returns that was documented in past work in the literature.

Our results can be summarised as follows. First, we find that PQM has had a positive effect on average test scores in mathematics but *not* in language. This effect is driven by large returns for stu-

[2] Normal age for enrolment in lower secondary education is 10, and progression to the upper secondary level - which is compulsory by the Italian law - is expected three years later. To give an example of how the *sezione* mechanism works, consider a school composed of 6 classes: 2 6th grade classes, 2 7th grade classes and 2 8th grade classes. This school will have 2 *sezione*, which we call A and B. Hence, in each year there will be *sezione* A class and *sezione* B class of 6th graders; *sezione* A class and *sezione* B class of 7th graders; and *sezione* A class and *sezione* B class of 8th graders. A student assigned to *sezione* A class in 6th grade in year 1, will be, with the same peers, in *sezione* A class in 7th grade in year 2, and so on.

dents of schools in the lowest tertile of pre-intervention achievement. These are learning environments characterised by students from markedly less advantaged backgrounds.

Second, for schools in the top tertile of pre-intervention achievement we find that extra hours tailored around reading activities have a *negative* effect on scores in mathematics, and no effect on language. This result is robust to a series of sensitivity checks that we perform in the data. Given that language abilities are found to be less responsive to PQM across learning environments, we conclude that in the least problematic environments instruction time should target activities that enhance mathematical abilities, as that the additional time spent at school engaged in language activities may substitute the time that students would have invested on mathematics.

Third, we find that PQM has increased inequality within the class. The counterfactual distributions of math scores (i.e., scores with and without PQM) differ only after the median. This implies that average effects of PQM are driven by a sharp change in the upper part of the score distribution within the class. Moreover, we find that the negative effect of language activities on mathematics for schools in the top tertile is concentrated at the top end of the test score distribution. In other words, extra time on language activities may come at the cost of outstanding performance in mathematics.

The remainder of the paper is organised as follows. Section 2 reviews the empirical literature on the effects of instruction time on academic achievement; in Section 3 we describe the intervention and provide some background information. Data are described in Section 4. Section 5 explains the sample selection criteria employed, and discusses descriptives statistics. In Section 6 we explain the methodology used, distinguishing between average and non linear (quantile) effects. Results and sensitivity checks are presented in Section 7 and Section 8, respectively, while Section 9 concludes by providing some policy recommendations.

2 Related literature

The rationale for increasing instruction time builds upon the direct effects of education on learning, and the side-benefits of lowered risk of negative behaviour for disadvantaged students. The empirical literature has adopted alternative strategies to assess the causal effect of instruction time on scores. A first strategy exploits between or within country variability in the exposure to school subjects. Small effects are documented in elementary and lower secondary schools by Lee and Barro (2001) using a panel of 59 countries, and by Wößmann (2003) using TIMMS data, in high schools by Lavy (2010) using PISA data. Rivkin and Schiman (2015) also use PISA data, and find that increased instruction time is more effective in better classroom environments. Mandel and Süssmuth (2011) exploit variation in instruction time across states in Germany and document positive effects on PISA text scores.

A second strategy exploits variation in length of school year resulting from quasi-experimental settings. Positive results for elementary and lower secondary schools students are documented in several studies: Marcotte (2007) and Marcotte and Hemelt (2008) consider number of school-closing days for snowfalls in Maryland, finding that students perform better in years with less unscheduled closing days. Hansen (2008) exploit weather-related cancellations in Colorado and Maryland, and changes in test-date administration in Minnesota. The results point to positive effects of the number of school days on student performance. Sims (2008) uses a similar idea exploiting a reform in Wisconsin, finding that increased time at school affects test scores in mathematics. Pischke (2007) exploits variation in instruction time resulting from the German "short school years", finding that shorter years are associated with higher grade repetition. Aucejo and Romano (2014) study the relative effectiveness of reducing absenteeism at school vis-à-vis extending the school calendar. Their findings indicate that the former strategy is most effective, in particular for low performing students. Similar effects are found for high school students, Bellei (2009) finds that the Chilean full school day programme has been beneficial for both reading and mathematics test scores, and Kikuchi (2014) exploits the revision of the Japanese curriculum standards in 1981 and shows that a 13 percent reduction of instruction time caused a 5 percent reduction in schooling and a 34 percent reduction in high school enrolment for women. Parinduri (2014) studies the effect of a reform implemented in all grades in Indonesia, which changed the start of school year from January to July. He finds that the longer school year decreases the probability of grade repetition and increases educational attainment, and it also increases the probability of working in formal sectors and wages later in life, especially for children living in rural areas.

A different strand of literature, which is closer in spirit to the intervention considered here, investigates the effects of increased school time conceived as "more hours per day" rather than "more days per year". Extra-education is organized by opening schools for longer hours during the afternoon, either providing extra instruction time on curricular activities or helping students from less advantage backgrounds doing their homework. Lavy and Schlosser (2005) consider a programme in Israeli high schools and document an increase in college matriculation rates of about 3 percent points. Jensen (2013) exploits a policy in Denmark that increased classroom hours in literacy and mathematics and documents very large returns in mathematics for 9th graders. Zimmer et al. (2010) considers Pittsburgh Public Schools and document positive effects for mathematics in both middle and elementary schools. Lavy (2012) exploits a policy experiment in Israeli elementary schools that changed the length of the school week and the time allocation to core subjects. He finds that more time at school on key tasks improves performance in mathematics, English and sciences, especially for students from low socio-economic backgrounds. On the contrary, Meyer and Van Klaveren (2013) find no significant effect on math or language achievement using data for Dutch elementary schools. Full school day compared to half school day was found to have a positive effect on learning outcomes also in kindergarten (Robin et al., 2006; DeCicca, 2007; Lash et al., 2008; Gibbs, 2010).

The literature on the effects of increased instruction time on academic performance overlaps, to a large extent, with that considering the effects of specific remedial programmes. Often students benefiting from increased time at school are from less advantaged backgrounds, and extra activities at school come in the form of remedial classes. The intervention considered here shares with remedial education programmes the idea that public investment should target schools at the lower end of the performance distribution. However, rather than targeting low achieving students in the school, PQM targets all students of low performing schools in the most deprived areas of the country. Positive effects of remedial education programmes are documented in Aiken et al. (1998), Calcagno and Long (2008), Bettinger and Long (2009) and De Paola and Scoppa (2014) for university students; in Lang et al. (2009) and Lavy and Schlosser (2005) for high school students. The experimental evaluation in Grossman and Sipe (1992) of a summer school programme aimed at contrasting early school dropouts finds no effects. Banerjee et al. (2007), using experimental data, study a programme providing remedial education to third and fourth graders in India. The intervention targets low achieving students in the treated schools, and offers remedial classes to small group of students during regular school time. They find an effect on test scores of about 0.40 points of standard deviations for children with the lowest pre-programme performance.

3 Background information on the intervention

Systematic evidence from international surveys (IEA TIMMS 2007; PISA 2003, 2006 and 2009) has identified a gap between Italy and other OECD countries. It is now well documented that Italian students perform below the European average in both mathematics and reading. This figure conceals a good deal of variability across regions, with Northern areas performing in line with other European countries and Southern areas performing markedly below. The recent experience on national assessment tests in Italy has demonstrated that, while the North/South divide is contained for second graders, it increases at the end of the primary school and grows even larger in middle schools (INVALSI, 2011). With the aim of boosting learning in lower and upper secondary schools, four regions located in the Objective 1 area (Campania, Sicily, Calabria and Apulia) became eligible for the EU Regional Development Funds and the European Social Fund for the period 2007-2013. PQM was one of the interventions funded through these resources.[3]

PQM targets *public* lower secondary schools in Objective 1 regions. It was first implemented in the year 2009/10, financing only additional instruction time in mathematics in 215 schools. In the following year new schools were added along with the possibility of extending instruction time to language. The total number of schools involved in the year 2010/11 was 223, 84 of which already participated in the first year. In either round, participation of schools was not mandatory. Applicant schools were enrolled giving preference to those under-performing with respect to percentage of repeating and failing students and dropout rates.[4] Non-random selection of schools, classes and, eventually, of intensity of afternoon activities are the three layers of potential bias challenging the empirical analysis.

Schools apply for PQM in June, and are notified with acceptance by the end of August. Families of children who will start sixth grade in September apply for school admission in February of the previous year, thus we can safely assume that prospective participation in PQM plays no role in parental choices regarding school enrolment. In addition mobility across schools - both before and after the school year has started - is limited because of administrative burdens and little negotiation power with the school. The transfer must be authorized by both principals at the schools of origin and destination; moreover, it must be motivated by objective reasons, usually a change of residence of the entire family. Importantly, participation of the school in PQM is communicated in September, not necessarily before class formation. We believe that hardly parents would change school because of PQM a few days before the academic year starts.

PQM schools must organise extra activities outside regular hours in a selected number of classes, two per subject. The school principal must name two teachers who will provide extra education at the time of application. If the school is enrolled in PQM, the choice made on teachers is strictly enforced. As participation in PQM should last for the three grades of lower secondary education, eligible teachers are solely those who with a permanent position and are expected to remain at the school for the full duration of the programme (ruling out, for example, teachers who are about to retire, or those employed with contracts with high turnover). Most likely these are teachers employed in the same *sezione* over time, and whose workloads at school are fulfilled in the sixth, seventh and eighth grade of the same *sezione* every year. Mobility of permanent staff is not necessarily zero, of course, but the cumbersome process required to transfer to a new school puts us on the safe side. The

[3] To the best of our knowledge, the only available evidence on the effects of the EU funds in Italy is Falzetti et al. (2012). By focusing on schools sampled in both the 2006 and the 2009 PISA waves, they find that schools in areas eligible to receive EU funds improved their test scores from 2006 to 2009 compared to schools belonging to regions of Southern Italy that are not eligible.

[4] We were not granted access to the list of applicant schools; because of this, our identification strategy cannot rely on exclusion restrictions defined using applicants denied participation in the programme. We know however that the pool of applicant schools that were eventually denied participation in 2010/11 is less than 15 percent (about 20 schools).

application for a transfer is filed by February and, if successful, is approved by June. Therefore we can safely assume that names of permanent teachers wishing to transfer are in the information set of principals at the time of application for PMQ. The salary offered to teachers for their extra loads is low (roughly 25 euro net per hour of activity), and excludes the time required to undertake training and the preparation of teaching materials. Because of this, we believe that PQM does not represent a strong monetary incentive to participation.

Participating teachers are mandated to a course that lasts 60 hours (30 of which are on-line), held during the first part of the academic year (October- December). The course does not target general competencies or knowledge of school subjects, but provides guidance on the organisation of afternoon activities, which are offered in the second part of the academic year (January-May). The course is held in groups of 10 teachers, and supervised by a mentor who provides support for the organisation of extra activities. Teachers are asked to draft an improvement plan based on a standardised test which student of participating classes take at the beginning of the academic year (end of September). The test should help teachers target pupils who are in need and the areas of intervention and in deciding how many and which kind of activities organize - remedial or advanced activities aimed at consolidating knowledge acquired during normal school hours.[5]

According to the programme regulation, the number of afternoon activities planned per class can vary between 1 and 8 and is agreed between principal and teachers once the students take the test, thus not at the time of application. Each activity consists of 15 hours of extra education to be held outside the regular school time to students, and the teacher can decide on intensity of exposure to these activities, as well as on contents (remedial education or advanced course). As a result, students receive an individualized education plan. In most classes the number of activities chosen by the teacher is between 2 and 4. Unfortunately we don't know to which activities students are assigned. However, qualitative information on the functioning of PQM suggests that the principle of equal participation to activities of all students in the class was the rule, rather than the exception. In other words, all students in PQM classes spent additional time at school. Because of this, our analysis is aimed at understanding the effects on scores of lengthening school hours.[6]

[5] We were not granted access to this test, and we cannot distinguish who in the class is benefitting from remedial or advanced activities.

[6] We focus on scores rather than on drop-out or failure rates, for two reasons: first, the population under study (sixth graders) hasn't yet reached the minimum age for leaving school (16 years). The most recent figures for drop-out rates at lower secondary school in the regions considered are all below 0.5% and our data, students aged 15 or more are 0.3%. Since school attendance is enforced by law until age 16, we don't think that dropout rates are important here. Second, failure rates would be interesting to consider, but unfortunately we do not have this kind of information. Indeed the test are taken roughly a month before the end of the school year and the data refer to this point in time, therefore we do not know whether the children subsequently failed or were passed to the next grade.

4 Data

School level data are provided by the Italian Ministry of Education through the INVALSI (*National Institute for the Evaluation of the Educational System*), and contain general information on characteristics (e.g. number of students, student to teacher ratio and dropout rates) and the exact municipality where the school is located. Through this information, geographical and demographic characteristics of the environment where the schools operate are also available.

Starting from the school year 2009/10, sixth grade students in Italy sit standardised tests under the supervision of the INVALSI. Participation of schools to the national test is compulsory by law. Students are tested in both mathematics and Italian language, and information is collected on socio-demographic characteristics (gender, year of birth, origin, level of education and employment status of the parents, household composition) as well as motivation and perception of the school.

Adopting the same international standards in IEA-TIMMS and PISA, INVALSI tests are designed to measure complementary dimensions of learning. Mathematics is assessed considering two cognitive domains: *knowledge* (which refers to the student's knowledge of facts, concepts, tools, and procedures in mathematics), and *reasoning* (which focuses on the student's ability to apply knowledge and conceptual understanding in a problem situation). Similarly, the test for language is designed to measure reading proficiency (i.e the ability to understand and interpret a text) and grammatical knowledge. We distinguish between outcomes that refer to Italian language (comprising *reading comprehension* and *grammatical knowledge*) and mathematics (comprising *mathematical knowledge* and *mathematical reasoning*). Scores used in the analysis represent the number of correct answers, and are standardized to have zero mean and unit variance by school year.

Past work has documented pervasive score manipulation in Southern regions of the country - see, for example, Angrist et al. (2014). These papers have looked at data for primary schools, where manipulation is acknowledged by the INVALSI and documented in their official publications. For lower secondary schools, however, the available documentation from INVALSI suggests that manipulation is less of a problem in the school grade and years considered in the analysis (INVALSI, 2010, 2011). Results from our analysis are unaffected if we control for (randomly assigned) external monitors in the school on the test day, as in Angrist et al. (2014).

It is finally worth being explicit about the sources of variability in the data used in what follows. Class is used to denote a group of students who are taught together at school. Cohorts are indexed to year of enrolment at school. Every cohort of students is divided into multiple classes at enrolment using Maimonides-like rules (Angrist et al., 2014). Classes are labelled with letters (e.g. A, B and C), and these denote *sezione* as we explained in the Introduction. The two relevant dimensions used to define our difference-in-differences comparisons are cohort of enrolment (2009 and 2010) and *sezione* in the school.

5 Sample selection and descriptive statistics

5.1 Sample selection criteria

We consider only the *second* wave of PQM, which was implemented in the year 2010/11. This choice is very pragmatic, and driven by the availability of standardised scores in Italy only from 2009/10. Our main sample selection consists in keeping only schools participating in PQM for the first time in the year 2010/11. Compared to all remaining schools on Objective 1 areas, PQM schools have more students, larger teacher-student ratios, more permanent teachers, and have applied for external funding in the past. However, past performance doesn't seem to be the main driver of school selection. The percentage of correct answers at the 2009/10 national test is 47.9 and 57.2 in PQM schools for math and language, respectively, while the corresponding figures in eligible regions are 48.2 and 57.2, respectively.

To contain differences due to voluntary participation, we preliminary matched PQM schools to non-participant schools with respect characteristics measured before 2010/11. We set out a *matched pair* comparison of similar schools located in the same province. Matching was implemented with replacement using region-specific propensity scores.[7] The working sample that resulted from this procedure is composed of 23 schools enrolled only in PQM mathematics, 37 schools enrolled only in PQM Italian language, and 74 schools enrolled in both components of PQM. This results in 127 classes receiving extra education in mathematics, 146 in Italian language and 40 in both subjects.

Using school identifiers provided by the INVALSI, we linked data for the same school in the 2009/10 (pre-programme) and the 2010/11 (post-programme) year. Importantly, we obtained identifiers for the group to which students are assigned at school (*sezione*). Thus our working sample consists of two consecutive cohorts of sixth graders enrolled in the same school and in the same *sezione* one year before and one year after the introduction of PQM. Table 1 presents the number of PQM and control schools, classes and students in both pre and post intervention years.[8]

[7]The matching procedure along the dimensions considered did not yield common support problems (only 4 PQM schools out of 138 were dropped because of this). Variables used for the calculation of the propensity score are: average percentage of correct answers in mathematics and language in sixth grade; student-teacher ratio, proportion of permanent teachers, dropout rate, failing rate, proportion of repeating students, proportion of immigrant students, proportion of disable students, proportion of female students, proportion of students attending more than 30 hours per week, number of students, whether the school has received in the previous year other PON funds for other activities, population in town and whether school is located on a mountain municipality. Only public schools were considered in the matching procedure.

[8]Figure A-1 of Appendix A shows a map of the 4 regions involved and the location and number of PQM and control schools in each municipality.

5.2 School characteristics and test scores

Table 2 presents descriptive statistics for PQM schools and for the schools chosen as controls in the working sample, using only pre-programme data. The average of the various dimensions considered is similar - see columns (1) and (2) - and, in fact, *not* statistical different between groups - see column (3). In column (4) we report the estimates of a logistic regression for the probability of being a PQM school in the working sample. It follows that none of the variables included is a good predictor for being a PQM school, and this corroborates the quality of the matching procedure implemented.[9]

Throughout our empirical exercise we will stratify schools in the matched sample according to tertiles of performance in mathematics in the pre-intervention year. The aim is to cluster schools according to their socio-economic background. A similar stratification could be obtained using test scores in Italian language, yielding conclusions similar to those presented below.[10] Summary statistics presented in Table 3 suggest that the stratification adopted indeed resembles division according to socio-economic background. As for student characteristics, we notice that students attending schools in the bottom tertile come from less-advantaged family backgrounds: less mothers are working, less fathers have a high occupational status, the proportion of parents with low education is much higher, and the proportion of parents with high education much lower; the index for home possession is lower.[11]

Table 4 paves the way for our empirical analysis, as it shows how test scores have changed over time in PQM and control schools. Descriptives are presented for mathematics and language by tertile of performance in the pre-programme year, the most problematic schools being in the top panel. As expected, test scores are highly correlated with the socio-economic background of students in the school. As for mathematics, the change in test scores for classes involved in PQM mathematics appears more pronounced, with respect to control classes, in the most problematic schools. For schools in the bottom tertile numbers in the table picture positive or constant trends in test scores from 2009/10 to 2010/11 for both mathematics and language; for schools in middle tertile trends are mixed; while scores in the best schools are always associate to negative trends.

[9] In Table A-1 of Appendix A we present descriptive statistics for student characteristics in the two groups of schools, that is for characteristics that were not directly used in the matching procedure. The table shows that there are minor differences between the two groups along some dimensions (e.g. students in control schools have higher percentage of mothers employed and of highly educated parents), but the overall results point to rather similar patterns.

[10] Main results obtained stratifying by Italian Language test score are reported in table A-3 in Appendix A.

[11] The variables used to calculate this index are: child has a quiet place to study; child has a desk to do his homework; child as a single room for herself; number of books in the house; house has an internet connection; house has a burglar alarm; house has more than one bathroom; parents have more than one car. Higher values of the score denote better off households.

5.3 Instruction time

Exploiting variation in number of activities provided, we can use *hours spent by the class in afternoon activities on school subject* (mathematics or language) as indicator of treatment intensity. In particular we considered a standardised version of this indicator, the percent change in instruction time obtained using the total number of hours that are mandatory dedicated to the subject during regular school time. We computed this by knowing that each activity lasts 15 hours, and we know that children in lower secondary schools dedicate 4 hours per week to mathematics and 7 hours per week to Italian language and that the number of weeks in a school year are 33. Descriptives of the indicator in the three groups of schools are provided in Table 5. The percentage change in instruction time is on average higher for mathematics than for language, and can be substantial: the mean value is well above 35 percent for mathematics, and set at about 25 percent for language (about 50 hours of afternoon activities for both subjects). The profile across school tertiles is hump-shaped, with schools at the two ends of the distribution of test scores presenting lower take up rates of extra activities.

To shed light into the possible determinants of this variability, we regressed our indicator of intensity in year 2010/11 on class inputs in year 2009/10, using observations for the same *sezione* in both years. The results of this analysis are fully documented in Table A-2 of Appendix A, and point to generally weak correlations of intensity with the regressors considered (i.e. test scores and student characteristics in the baseline year).

6 Methods

6.1 Identification

Selection of classes into PQM is modeled using *sezione* fixed effects. Schools can autonomously set the rules for the assignment of teachers to *sezione*, but the principle of ensuring that students maintain the same teacher over the secondary education cycle (three years, from grade six to grade eight) is strongly recommended by the Italian law. Teachers work in a highly regulated public sector, with virtually no risk of termination, and are subject to a pay and promotion structure that is largely independent of performance. Teachers hired on a permanent contract - a requirement to participate in PQM - are employed at school 18 hours per week. As instruction time in mathematics is set by law to 6 hours per week in each grade of lower secondary education, the practice that math teachers fulfil duties over the three grades of the same *sezione* is widespread (the 6 hours include 2 hours devoted to science). This rule is also prompted by a national reform on staff ratios that passed before the beginning of PQM (*Decreto ministeriale number 37, 26 March 2009*), stating that the number of math teachers employed at school must correspond to the number of *sezione* activated. In this setting, controlling for *sezione* fixed effects most likely controls for math teacher effects.

As for Italian language, instruction time is set to 9 hours per week (one hour is devoted to history and geography). This implies that two language teachers are usually assigned to the three grades of the same *sezione*. Indeed, the above mentioned reform states that the head count for language teachers is computed as 1.5 times the number of *sezione* activated. For practical matters, it is common to rotate teachers across grades, thus limiting their mobility across *sezione* over time.[12] In this setting, *sezione* fixed effects arguably model the composition of teachers of the class.

The analysis is carried out using two cohorts of sixth graders in 2009/10 and 2010/11. All outcomes are defined at the class level. We compare performance of students enrolled in the same *sezione* before and after PQM. This defines a standard difference-in-differences approach, with *sezione* fixed effects. By controlling for *sezione* fixed effects we indirectly control for school fixed effects, and thus for all sources of potential biases related to unobservable characteristics of the class and the school. The causal parameters retrieved refer to the effect of being in a PQM class.

Two threats to the validity of the strategy discussed are the possible direct effects of PQM on mobility of principals (and teachers), as well as on *sezione* composition over time. We believe that the former shouldn't be of concern here, given the labor market of teachers and principals in Italian public schools. The procedure required for a transfer, combined with the main features of PQM described above are arguably weak determinants of mobility. On the other hand, we checked if the introduction of PQM affected *sezione* composition. Results not presented here, and obtained using the same equations in the next section, rules out changes to class formation along key demographics available in the data (gender, ethnicity, family background and parental education).

6.2 Estimation

Our analysis allows for cross subject effects (i.e. effect of PQM mathematics on test scores in language, and *viceversa*). Our preferred specification considers the following equation which is estimated by school subject:

$$y_{jtk} = \alpha_{jk} + \beta_k^M D_{jt}^M + \beta_k^I D_{jt}^I + \delta_k N_{jt} + \eta_{tk} + \epsilon_{jtk}, \quad (1)$$

where y_{jtk} is the outcome variable in *sezione* j, year t ($t = 1$ and $t = 2$ refer to the pre- and post-programme periods, respectively) and school subject k (mathematics and Italian language). D_{jt}^M and D_{jt}^I are dummies for being enrolled in *any* activity in mathematics (M) and Italian language (I), respectively, while N_{jt} is dummy for control classes in PQM schools. There is $D_{j1}^I = D_{j1}^M = N_{j1} = 0$

[12]Given the lack of administrative data, no empirical evidence can be provided to support the adoption of such practice. However anecdotal evidence - also confirmed by the Ministry of Education - suggests that this is the predominant rule rather than the exception.

for all j's. With this notation β_M^M and β_M^I measure the effect of receiving extra hours in mathematics or language, respectively, on mathematics scores. A similar interpretation applies to the effects β_I^M and β_I^I on language; η_{tk} captures time effects and α_{jk} is the *sezione* fixed effect, both subject specific, and ϵ_{jtk} is a random error.

Equation (1) is then refined by considering variability in length of afternoon activities. The following specification is considered:

$$y_{jtk} = \alpha_{jk} + \beta_k^M D_{jt}^M + \lambda_k^M H_{jt}^M + \beta_k^I D_{jt}^I + \lambda_k^I H_{jt}^I + \delta_k N_{jt} + \eta_{tk} + \epsilon_{jtk}, \tag{2}$$

where H_{jt}^M and H_{jt}^I represent intensity in mathematics and language, respectively, as deviation from the mean. The parametrization adopted is such that the coefficients β_k^M and β_k^I in equation (2) can be read as the effect of PQM when intensity is set to its mean, which is about 50 hours for both subjects. As the length of afternoon activities is chosen by teachers, for identification we rely on *sezione* fixed effects.

Standard errors are clustered at the school level. As a sensitivity check we clustered standard errors at the *sezione* level, which is where treatment takes place. This analysis yields results which are informational equivalent to those presented below. The conditioning on a set X_{jt} of student and school level variables is left implicit throughout, but is used in the empirical analysis. All regressions presented control for class size, weekly class schedule and class level variables constructed from student level information (e.g. gender, immigration status, whether the student is ahead or behind compared to her age, education of parents and mother working status, plus control for missing values).

The analysis is carried out by considering *three* different outcomes y_{jtk}, all defined at the class level. We start by using average scores, for which results are presented in Table 6. We then set y_{jtk} to the percentage of students in the class scoring above a certain threshold, which we make subject specific. Thresholds are defined using baseline data, calculating a grid from the 1^{st} to the 99^{th} percentile of the relevant distribution. By considering this outcome variable, we use *within class variability* in scores to study possible heterogeneity in the effects of PQM. Results using this outcome are in Figure 1. Our approach closely resembles a non-linear difference-in-differences method, that was originally proposed by Firpo et al. (2009) and then re-considered by Havnes and Mogstad (2010). Our specification for dealing with quantiles proves particularly convenient to account for the availability of multiple control groups, and to model treatment intensity represented by variability across PQM classes in the number of activities. Finally we use between class variability in scores to study heterogeneity in the effects of PQM across classes. To this end, we consider an indicator for having the class average score above a certain threshold, which again we make subject specific and defined from percentiles calculated on baseline data. The results from this analysis are presented in Figure 2.[13]

7 Results

7.1 Average effects

Table 6 presents estimates of the effects obtained from equation (1) for both mathematics and language, by tertiles of school performance in the pre-intervention year. The variable considered is average score in the class. The left hand side part of the table refers to mathematics, while the right hand side part refers to language. For both school subjects we present the break down by cognitive domains covered in the test.

The first result worth noting is the absence of any effect of PQM on language, for both domains considered and across tertile groups - see columns (4), (5) and (6). On the other hand, we find that extra instruction time in mathematics has large positive effects on mathematics, but only for students in the most problematic schools (0.29 points) - see column (1). This effect is driven by the "reasoning" domain; the effect on the "knowledge" domain is positive, but less statistically significant and of slightly smaller magnitude.[14] This finding suggests that extra instruction time increase basic knowledge and helps students apply and use knowledge acquired during normal school hours.

We find evidence of cross subject effects for schools in the top tertile, since extra time dedicated to language impacts negatively on scores in mathematics - see column (3). The size of this effect (-0.28 points) is comparable to that of extra activities in mathematics found for schools in the bottom tertile, and it is mostly driven by the "knowledge" domain. This result suggests that the extra time spent on activities targeting Italian language substitutes the time that the students in best schools would have spent developing their mathematical knowledge, and this in turn impacts on scores. The general picture that emerges is in favour of targeting the most problematic students with extra activities that help develop mathematics skills. Moreover, we find that test scores for control classes are the same in schools with and without PQM - this is the coefficient on N_{jt} in equation (1). This rules out possible spill-over effects that may arise, for example, if teachers of PQM classes share with other colleagues at the school materials and teaching modalities acquired during the preparatory course.

Finally, Table 7 replicates the same analysis by considering equation (2). Results for the average effects are reported in columns (1), (5) and (9). Coefficients for participation into any mathematics

[13] We performed various sensitivity checks finding that results are robust to modifications of the specification adopted. In particular, we estimated quantile treatment effects relying on the change-in-changes approach by Athey and Imbens (2006). Alternative estimation strategies yielded results similar to those presented below. Selected results of the change-in-changes are presented in table A-4 of Appendix A.

[14] A bootstrap test for the equality of results for knowledge and reasoning fails to reject the null at the conventional levels.

and Italian language activities confirm findings from equation (1); in addition it seems the length of additional instruction time does not seem to play any role, since the variation from the mean it is never significantly different from zero, with the exception of schools in the top tertile which increase slightly test scores in Italian if they receive more classes than the average in Italian and math - column (9).

Given the number of outcome variables considered, we run the risk of over rejection of the null hypothesis due to problem of multiple inference. To assess the robustness of our results to this problem, we implement the the procedure by Westfall and Young (1993) to adjust p-values. Their step-down permutation algorithm is applied to the two main coefficients in Table 6 by tertile (any extra class in mathematics and any extra class in language). Considering the following scores: math, knowledge, reasoning, language, grammar and reading. As the procedure by Westfall and Young (1993) is often considered too conservative, we also check the sensitivity of conclusions using two alternative methods. First, we obtain p-values from a non-parametric permutation test (Anderson, 2008; Efron and Tibshirani, 1993). Second, we compute false discovery rate (FDR)adjusted p-values, as presented in Anderson (2008). The adjusted p-values substantially confirm the results obtained in Table 6 (see Table A-5).

7.2 Quantile effects

We start by defining a grid that ranges between the 1^{st} and the 99^{th} percentiles of the distribution of scores in the baseline period, for mathematics and language. We then consider the fraction of students in the class whose score is above each percentile. This defines 99 variables, which we employed as outcomes in equation (1). Figure 1 provides a graphical summary of estimation results (dashed lines refer to 95 percent confidence intervals). Graphs on the left hand side report the effect of PQM-mathematics on the fraction of students with a math score above each percentile. Similarly, graphs on the right hand side report the effect of PQM-language on the fraction of students with a language score above each percentile. The top, central and bottom panels of the figure refer to schools in the bottom, medium and top tertile of pre-intervention scores, respectively. The discussion on cross-subject effects is deferred to Figure 3.

Consistently with the first row of Table 6, results in Figure 1 are significant only for mathematics in the most disadvantaged schools. In these environments PQM increased inequality within the class: high math scores (i.e. scores above the median) are more likely, but the occurrence of low scores is not affected by the intervention. This effect on the distribution of math scores boosts average performance of the class in the most disadvantaged schools, which explains the effect in column (1) of Table 1.

However, Figure 2 shows that not all classes have the same shift in average performance. The figure is constructed as Figure 1, but here the outcome in equation (1) is a dummy that takes value one when the average score in the class is above each percentile. In other words, Figure 1 considers *within* class variability whereas Figure 2 looks at *between* class variability. In the most disadvantaged schools, the bottom 40 percent of the score distributions in PQM and non-PQM classes coincide. In results non presented, we find that a similar picture emerges for mathematical reasoning. These findings are consistent with the idea that PQM has had no effect in the worst learning environments.[15]

Figure 3 reports the same quantile analysis for the effects of PQM-language on math scores considering schools in the top tertile. This is the only cross subject effect for which we find statistically significant results in column (3) of Table 6. The graph on the left hand side of is obtained as Figure 1; the graph on the right hand side instead replicates the analysis in Figure 2. We find that the negative effect documented in the Table 6 is substantially stable across classes, although it is very imprecisely measured. Such effect appears to be driven by negative effects at the highest quantiles of the score distribution within the class. This finding is consistent with the idea that the extra activities targeting mastery of Italian language result in lower performance in mathematics for the best students.

Columns (2)-(4), (6)-(8) and (10)-(12) of Table 7 replicate the *within class* analysis by considering our measure of treatment intensity exposure. Only results for selected percentiles (25^{th}, 50^{th} and 75^{th}) are presented. As for the effects on mathematics test scores - panel A- for schools in the bottom tertile, the general pattern in Figure 1 is confirmed, with larger effects documented in top quantiles - columns (3) and (4). Similarly, for schools in middle tertile we find no effects. For schools in the top tertile we find that the duration of afternoon activities plays some role at the bottom end of the distribution (see results for the bottom 25 and 50 percent- column (10) and (11)). Also confirmed are the negative cross subject effects of activities in Italian language on mathematics test scores, which are increasing in size as one moves to the highest percentiles of the distribution (columns (11) and (12)). We seem to find evidence that longer hours spent on individualised instruction become important only if we consider students from the best performing schools. We also find some effects on the Italian language test score (panel B): in particular for schools in the top tertile who receive more classes than the average in Italian language and mathematics. And for school in the middle tertile we find a negative cross subject effects of receiving more classes in mathematics on Italian language.

[15] Additional assumptions (e.g., rank invariance) are required to interpret our results as student-level or class-level treatment effects.

8 Robustness checks

The aim of this section is twofold. First, we then discuss the plausibility of the common trend assumption for PQM and control schools in pre-intervention periods. Since tests scores from the national assessment are available in Italy for sixth graders only from 2009/10, we proxy counterfactual scores by considering fifth graders, for whom data are available since 2008/09. Second, we provide a test for the validity of *sezione* fixed effects as proxy for unobserved class differences in our analysis.

8.1 Specification test using test scores for fifth graders

The assumption needed for our identification strategy to work is that, in the absence of PQM, scores in all classes would have presented parallel trends. Scores for sixth graders are available at the national level starting from the year 2009/10. This makes it impossible to test for the existence of pre-intervention trends in the outcomes of interest using only our working sample. For this reason, we use scores for fifth graders that are available at the national level from 2008/09, thus adding one year of pre-intervention data.[16] The key assumption is that the trend in scores for fifth graders in the years before PQM provides a good approximation to the trend for sixth graders. This assumption is rather innocuous in the Italian context, given that the transition from the primary (fifth grade) to the lower secondary (sixth grade) school is characterised by extremely low geographic mobility. Students typically enrol at both levels in schools in the same *local* area, which in the large majority of cases are only a few kilometers apart or are located in adjacent buildings. This means that many of the students who complete primary education in June, end up enrolled in sixth grade at the same school in September.

We therefore selected fifth graders enrolled in primary schools of areas where PQM was implemented, and used this group to proxy pre-programme trends from 2008/09. As the sixth grade is the first year at lower secondary school, we need to establish a link between PQM schools in our sample and primary schools of Objective 1 areas. We proceed according to the following steps. First, we define a dummy "PQM municipality" depending on number of lower secondary schools enrolled in PQM. We consider *four* alternative definitions as a sensitivity check for our conclusions. The first definition imposes that a PQM municipality must have *all* lower secondary schools enrolled in PQM from 2010/11. Clearly, this definition may be too restrictive as - for example - larger municipalities are most likely excluded because they have a larger number of schools. The alternative definitions are less restrictive, and impose that PQM municipalities must have at least 20 and 50 percent of

[16] In 2008/09 participation was based on voluntary bases. Nevertheless a random sample of schools - about 25 percent of schools in the population- was mandated to participate and test scores were administered in presence of an external monitor. We only use this sample of schools for 2008/09.

lower secondary schools enrolled in PQM from 2010/11. The last definition that we employ is the least restrictive, and labels PQM municipalities as those where at least one school is involved in PQM from 2010/11. Second, we collapse data at the municipality level, and compute average test scores for students enrolled in fifth grade at primary school. We then compare the time series of test scores for fifth graders in PQM municipalities to that of fifth graders in non-PQM municipalities, for years 2008/09, 2009/10 and 2010/11. Figure 4 presents the times series of fifth graders for the four definitions of PQM municipality adopted. We present results only for math scores, as the analysis yielded similar conclusions for language. Regardless of the definition adopted, the time series for fifth graders present remarkably similar trends over time, thus corroborating the assumption needed for identification. Under the assumptions stated, we can conclude that sixth grade students in PQM areas have competencies at entrance that are similar to those of students in other municipalities that in the main analysis are used as controls.

8.2 Class vis-à-vis *sezione* clustering of students

We start by estimating a model in which the variance of test scores across students is explained by two levels: school and class. In this estimation we consider all schools located in the four Objective 1 regions, excluding PQM schools. We find that, for both subjects, the fraction of variance explained by each level is roughly constant in 2009/10 and 2010/11. We thus pool the two waves of data, and report in Panel A of Table 8 results from the multilevel analysis. Class membership is twice as much important than school membership in explaining variability of mathematics scores (21 percent as opposed to 11 percent), while it is less for language scores. We also report results of the same analysis obtained using *sezione* membership instead of class membership. Since the same *sezione* appears both in 2009/10 and 2010/11, this analysis attempts at explaining the same variance using roughly half the number of groups. We find that the fraction of variance explained by school membership is roughly comparable to the previous figures, and that *sezione* explains $12/21 = 57$ (for mathematics) and $9/14 = 64$ (for language) percent of the variance that was explained using class membership. We conclude that the contribution to the variance coming from class membership is mostly spanned by that coming from *sezione* membership.

We complement this analysis by regressing average test scores in 2010/11 on average test scores in 2009/10 for the same *sezione*. Three nested specifications are considered for the regressions reported in Panel B of Table 8. The first specification obtains results from a null model; the second controls for school fixed effects; the third adds controls (percentage of females, class size, percentage of foreign students). The following outcomes are considered: test scores (to proxy persistence of students' and teachers' ability), and the home possession index and the percentage of foreign students in the class (to proxy assignment based on socio-economic status). Results show that the persistence in the characteristics of the *sezione* cannot be fully explained by school fixed effects or observable variables.

9 Conclusions

This paper has provided evidence on the effect of mandating students from low achieving schools in Southern Italy to afternoon activities in mathematics and Italian language. We have investigated the effects of the Quality and Merit Project (PQM), implemented in Objective 1 regions through EU Regional Development Funds and EU Social Funds. We have found that increasing time spent at school is important only in the most problematic learning environments. In our setting, these are represented by schools in the bottom tertile of the distribution of performance in the pre-PQM period.

We have found that only scores in mathematics are affected, and estimated a positive effect of about 0.296 points of standard deviation. In line with the objectives of the intervention, we have shown that this result follows from an improvement in quantitative *reasoning* and in mathematical *knowledge*. This result is consistent with other studies in the literature showing that it is much harder to intervene on reading and comprehension skills, rather than on skills involving practice, like maths, because a large part of literacy work takes place through general vocabulary training in the home environment (Marcotte, 2007; Sims, 2008; Zimmer et al., 2010; Jensen, 2013). Research in developmental psychology has suggested that the critical period for language development occurs early in life, while the critical period for developing higher cognitive functions extends into adolescence, thus as it is documented in previous works that considers the impact of early interventions on children outcomes, in order to affect reading skills and language test scores we should target younger children, during elementary or pre-schools.

We have concluded that the first order effects of being assigned to individualised activities are more important than those arising from the duration of these activities. We also have found that the positive effect of PQM in the most critical schools conceals a good deal of variability *within* the class. We have shown that the effect is driven by a large, positive shift to the distribution of test scores, but that a sizeable group at the bottom end of the distribution is left unaffected by the intervention. As in other studies that have investigated similar interventions - see, for example, Lavy and Schlosser (2005) and Banerjee et al. (2007) - we have found that only the least advantaged students are those with positive returns to participation.

Our results imply that EU Regional Development Funds and EU Social Funds used to roll out education policies in the most deprived areas must tailor effective strategies around the students most in need to avoid important deadweight loss.

Acknowledgments

This paper benefited from discussions with Daniele Checchi and Enrico Rettore, and from comments by audiences at UPF (October 2012), Dondena, Bocconi University (December 2012), the Fifth Italian Congress of Econometrics and Empirical Economics (January 2013), University of Turin, (February 2013), IV Workshop on Public Policies, Social Dynamics and Population (March 2013) and Queen Mary University of London (April 2013). We thank the INVALSI for providing the data and in particular Daniele Vidoni, Patrizia Falzetti, Michele Cardone and Andrea Caputo for continuous support and INDIRE, in particular Samuele Calzone, Sara Mori and Serena Greco. Part of the work was carried out at the JRCs Centre for Research on Education and Lifelong Learning (CRELL) in the framework of the project KNOW (Human Capital for Prosperity and Sustainable Growth in Europe). We thank four anonymous referees for helpful comments and suggestions.

References

Aiken, L. S., S. G. West, D. E. Schwalm, J. L. Carroll, and S. Hsiung (1998). Comparison of a randomized and two quasi-experimental designs in a single outcome evaluation efficacy of a university-level remedial writing program. *Evaluation Review 22*(2), 207–244.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association 103*(484), 1481–1495.

Angrist, J. D., E. Battistin, and D. Vuri (2014). In a small moment: class size and moral hazard in the mezzogiorno. Technical report, No. 20173. National Bureau of Economic Research.

Athey, S. and G. W. Imbens (2006). Identification and inference in nonlinear difference-in-difference models. *Econometrica 74*(2), 431–497.

Aucejo, E. M. and T. F. Romano (2014). Assessing the effect of school days and absences on test score performance. Technical report, CEP Discussion Papers, CEPDP1302. Centre for Economic Performance, London School of Economics and Political Science, London, UK.

Banerjee, A. V., S. Cole, E. Duflo, and L. Linden (2007). Remedying education: Evidence from two randomized experiments in india. *The Quarterly Journal of Economics 122*(3), 1235–1264.

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

Bettinger, E. P. and B. T. Long (2009). Addressing the needs of underprepared students in higher education does college remediation work? *Journal of Human Resources 44*(3), 736–771.

Calcagno, J. and B. Long (2008). The impact of postsecondary remediation using a regression discontinuity approach: Addressing endogenous sorting and noncompliance. Technical report, No. 14194. National Bureau of Economic Research.

De Paola, M. and V. Scoppa (2014). The effectiveness of remedial courses in italy: a fuzzy regression discontinuity design. *Journal of Population Economics 27*(2), 365–386.

DeCicca, P. (2007). Does full-day kindergarten matter? evidence from the first two years of schooling. *Economics of Education Review 26*(1), 67–82.

Efron, B. and R. J. Tibshirani (1993). An introduction to the bootstrap, monographs on statistics and applied probability, vol. 57. *New York and London: Chapman and Hall/CRC*.

Falzetti, P., L. Palmerio, B. Romano, and D. Vidoni (2012). Reducing the gap? the effects of eu funds for improving education in southern italy. Vancouver, British Columbia, Canada. AERA Annual meeting.

Firpo, S., N. Fortin, and T. Lemieux (2009). Unconditional quantile regressions. *Econometrica 77*(3), 953–973.

Gibbs, C. (2010). Measuring the impact of full-day kindergarten: Experimental and quasi-experimental evidence. *Society for Research on Educational Effectiveness*.

Grossman, J. B. and C. L. Sipe (1992). *Summer Training and Education Program (STEP): Report on long-term impacts*. Public/Private Ventures.

Hansen, B. (2008). School year length and student performance: Quasi-experimental evidence. *Unpublished manuscript. University of California, Santa Barbara*.

Havnes, T. and M. Mogstad (2010). Is universal child care leveling the playing field?: Evidence from non-linear difference-in-differences. *IZA Discussion paper*.

INVALSI (2010). Rilevazione degli apprendimenti snv 2009/10. prime analisi. Technical report, INVALSI.

INVALSI (2011). Servizio nazionale di valutazione 2010-11. le rilevazioni degli apprendimenti a.s. 2010-11. Technical report, INVALSI.

Jensen, V. M. (2013). Working longer makes students stronger? the effects of ninth grade classroom hours on ninth grade student performance. *Educational Research 55*(2), 180–194.

Kikuchi, N. (2014). The effect of instructional time reduction on educational attainment: Evidence from the japanese curriculum standards revision. *Journal of the Japanese and International Economies 32*, 17–41.

Lang, L., J. Torgesen, W. Vogel, C. Chanter, E. Lefsky, and Y. Petscher (2009). Exploring the relative effectiveness of reading interventions for high school students. *Journal of Research on Educational Effectiveness 2*(2), 149–175.

Lash, A., S. Bae, V. Barrat, E. Burr, and T. Fong (2008). Full-day kindergarten and student achievement: A literature review. Technical report, US Department of Education, REL West Technical Assistance Response. Washington, DC.

Lavy, V. (2010). Do differences in schools instruction time explain international achievement gaps in math, science, and reading? evidence from developed and developing countries. Technical report, No. 16227. National Bureau of Economic Research.

Lavy, V. (2012). Expanding school resources and increasing time on task: Effects of a policy experiment in israel on student academic achievement and behavior. Technical report, No. 18369. National Bureau of Economic Research.

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

Lee, J. and R. Barro (2001). Schooling quality in a cross–section of countries. *Economica 68*(272), 465–488.

Mandel, P. and B. Süssmuth (2011). Total instructional time exposure and student achievement: An extreme bounds analysis based on german state-level variation. *CESifo Working Paper Series*.

Marcotte, D. (2007). Schooling and test scores: A mother-natural experiment. *Economics of Education Review 26*(5), 629–640.

Marcotte, D. and S. Hemelt (2008). Unscheduled school closings and student performance. *Education Finance and Policy 3*(3), 316–338.

Meroni, E. C. (2013). *Average and quantile effects of more instruction time in low achieving schools: evidence from Southern Italy.* Ph. D. thesis, Department of Statistical Science, University of Padua.

Meyer, E. and C. Van Klaveren (2013). The effectiveness of extended day programs: Evidence from a randomized field experiment in the netherlands. *Economics of Education Review 36*, 1–11.

Parinduri, R. A. (2014). Do children spend too much time in schools? evidence from a longer school year in indonesia. *Economics of Education Review 41*, 89 – 104.

Pischke, J.-S. (2007). 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.

Rivkin, S. G. and J. C. Schiman (2015). Instruction time, classroom quality, and academic achievement. *The Economic Journal 125*(588), 425–448.

Robin, K., E. Frede, and W. Barnett (2006). Is more better? the effects of full-day vs half-day preschool on early school achievement. *NIEER Working Paper*.

Sims, D. (2008). Strategic responses to school accountability measures: It's all in the timing. *Economics of Education Review 27*(1), 58–68.

Westfall, P. H. and S. S. Young (1993). *Resampling-based multiple testing: Examples and methods for p-value adjustment*, Volume 279. John Wiley & Sons.

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

Zimmer, R., L. Hamilton, and R. Christina (2010). After-school tutoring in the context of no child left behind: Effectiveness of two programs in the pittsburgh public schools. *Economics of education Review 29*(1), 18–28.

Table 1: Sample size (schools, classes and students)

		Pre treatment year	Post treatment year
Schools	Enrolled in PQM	134	134
	Used as control	114	114
Classes	Enrolled in PQM	313	313
	Used as controls in PQM schools	407	407
	Used as controls in remaining schools	595	595
Students	Enrolled in PQM	6,228	6,461
	Used as controls in PQM schools	8,260	8,380
	Used as controls in remaining schools	12,455	12,672

Note. Presented are number of students, classes and schools in the working sample (see Section 5.1 for details).

Table 2: Descriptives at the school level of pqm and control schools

	(1) PQM	(2) Control	(3) Difference	(4) Score
% of correct answers in mathematics, sixth grade	0.479	0.490	-0.010	1.367
			(0.010)	(3.045)
% of correct answers in language, sixth grade	0.572	0.585	-0.014	-5.041
			(0.007)	(3.860)
Proportion of teachers with permanent contract	0.892	0.904	-0.012	-0.261
			(0.012)	(1.609)
Student to teacher ratio	9.632	9.931	-0.299	-0.097
			(0.275)	(0.100)
Number of students	402.8	398.4	4.470	0.001
			(26.34)	(0.001)
Proportion of foreign students	0.027	0.027	0.000	1.379
			(0.003)	(5.129)
Proportion of students with disabilities	0.034	0.031	0.003	2.345
			(0.003)	(8.652)
Drop out rate	0.003	0.003	0.000	-5.394
			(0.001)	(12.208)
Failure rate	0.049	0.046	0.003	-3.374
			(0.006)	(3.832)
Proportion of repeating students	0.048	0.041	0.007	3.453
			(0.006)	(4.130)
Proportion of females in the school	0.490	0.488	0.002	0.557
			(0.007)	(2.395)
Proportion of classes doing more than 30 hours	0.335	0.337	-0.001	-0.170
			(0.051)	(0.401)
School received PON funds for students'activities	0.963	0.974	-0.011	-0.239
			(0.023)	(0.770)
Municipality located on montain	0.284	0.246	0.038	0.179
			(0.056)	(0.307)
(Log) population in town, 2009	10.38	10.31	0.069	0.004
			(0.192)	(0.111)
Number of schools	134	114		

Note. Presented are descriptive statistics for schools in the working sample (pre-programme data only). Column (1): schools enrolled; column (2): control schools; column (3): difference between column (1) and column (2) (standard error in parentheses); column (4): logit regression for being a PQM school (standard errors in parentheses).

Table 3: Descriptive statistics (schools by tertiles of 2009/10 performance in mathematics)

	Tertiles of Test Scores		
	Bottom	Middle	Top
Test score mathematics	-0.478	-0.014	0.439
Test score language	-0.376	-0.024	0.282
Proportion of teachers with permanent contract	0.874	0.892	0.926
Student to teacher ratio	9.407	10.143	9.758
Number of students	366.7	427.2	408.4
Proportion of females in the school	0.487	0.483	0.498
Proportion of foreign students	0.029	0.023	0.029
Proportion of students with disabilities	0.040	0.029	0.029
Drop out rate	0.006	0.001	0.002
Failure rate	0.064	0.044	0.035
Proportion of repeating students	0.066	0.039	0.030
Class weekly hour	31.70	31.18	31.91
Class size	21.84	22.33	22.35
Proportion of students whose parents have low education	0.534	0.426	0.350
Proportion of students whose parents have medium education	0.360	0.409	0.449
Proportion of students whose parents have high education	0.106	0.165	0.201
Proportion of mothers employed	0.338	0.397	0.432
Proportion of students whose father's occupation is : unemployed	0.078	0.061	0.049
Proportion of students whose father's occupation is : blue collar	0.337	0.308	0.274
Proportion of students whose father's occupation is : white collar	0.432	0.415	0.429
Proportion of students whose father's occupation is : managerial	0.153	0.216	0.249
Proportion of students living with both parents	0.890	0.896	0.903
Average HOME scale coefficient	-0.168	-0.057	0.048
Number of schools	82	82	84

Note. Presented are descriptive statistics for all schools in the working sample. Tertiles are defined from the distribution of test scores in mathematics in year 2009/10 (see Section 5.2 for details).

Table 4: Descriptive statistics (changes over time in test score)

	Mathematics			Language		
Bottom tertile	Pre	Post	Change	Pre	Post	Change
Control class in control schools	-0.526	-0.459	0.154	-0.372	-0.325	0.134
Control class in PQM schools	-0.474	-0.359	0.166	-0.308	-0.281	0.079
Any extra class in Language	-0.458	-0.354	0.103	-0.355	-0.296	0.058
Any extra class in Mathematics	-0.487	-0.432	0.419	-0.068	-0.217	0.215
	Mathematics			Language		
Middle tertile	Pre	Post	Change	Pre	Post	Change
Control class in control schools	-0.031	0.036	0.05	0.019	0.059	0.023
Control class in PQM schools	-0.045	-0.016	-0.039	-0.083	-0.058	-0.042
Any extra class in Language	0.085	0.002	-0.112	-0.026	0.012	0.010
Any extra class in Mathematics	-0.017	-0.120	0.063	0.046	-0.025	0.095
	Mathematics			Language		
Top tertile	Pre	Post	Change	Pre	Post	Change
Control class in control schools	0.373	0.264	-0.172	0.201	0.143	-0.121
Control class in PQM schools	0.353	0.271	-0.175	0.179	0.101	-0.169
Any extra class in Language	0.504	0.32	-0.269	0.236	0.293	-0.027
Any extra class in Mathematics	0.400	0.189	-0.208	0.192	0.093	-0.096

Note. Reported are average test scores by group calculated for 2009/10 (pre-programme) and 2010/11 (post-programme) data. Reported also is the change over time of test scores. Test scores have been standardised to have zero mean and unit variance in each year and subject. Schools are grouped into tertiles defined from the average test score in mathematics in year 2009/10 (See Section 5.2 for details).

Table 5: Descriptive statistics (exposure to afternoon activities)

		Bottom tertile	Middle tertile	Top Tertile
Mathematics				
	25^{th}	22.72	34.09	22.72
	50^{th}	34.09	45.45	34.09
Percent change in instruction time	Mean	36.83	44.65	36.54
	75^{th}	45.45	45.45	45.45
Italian language				
	25^{th}	19.48	19.48	19.48
	50^{th}	19.48	25.97	19.48
Percent change in instruction time	Mean	22.64	26.96	21.55
	75^{th}	25.97	32.47	25.97

Note. Presented are descriptive statistics for the measure of exposure to afternoon classes in mathematics and Italian language for the three groups of schools. See Section 5.3 for details.

Table 6: Effect of PQM on mathematics and Italian language (participation vis-à-vis non participation)

	(1) Bottom tertile	(2) Middle tertile	(3) Top tertile	(4) Bottom tertile	(5) Middle tertile	(6) Top tertile
	Mathematics			Italian language		
Any extra class in mathematics	0.296***	0.054	0.030	0.020	0.061	0.086
	(0.112)	(0.092)	(0.088)	(0.094)	(0.071)	(0.097)
Any extra class in language	-0.004	-0.059	-0.284**	-0.103	0.050	-0.021
	(0.084)	(0.084)	(0.119)	(0.083)	(0.082)	(0.105)
Control class in PQM schools	0.015	0.037	-0.033	-0.056	0.017	-0.063
	(0.069)	(0.071)	(0.089)	(0.069)	(0.054)	(0.086)
	Mathematical reasoning			Reading comprehension		
Any extra class in mathematics	0.308***	0.095	-0.007	0.000	0.118	0.029
	(0.090)	(0.091)	(0.098)	(0.083)	(0.075)	(0.097)
Any extra class in language	-0.012	-0.040	-0.217*	-0.086	0.069	-0.047
	(0.080)	(0.088)	(0.119)	(0.074)	(0.076)	(0.107)
Control class in PQM schools	0.047	0.010	-0.008	-0.085	0.020	-0.073
	(0.067)	(0.080)	(0.088)	(0.062)	(0.058)	(0.081)
	Mathematical knowledge			Grammar knowledge		
Any extra class in mathematics	0.247**	0.025	0.059	0.090	-0.034	0.089
	(0.124)	(0.088)	(0.080)	(0.101)	(0.083)	(0.090)
Any extra class in language	-0.005	-0.071	-0.302***	-0.054	0.021	0.011
	(0.083)	(0.078)	(0.114)	(0.090)	(0.085)	(0.096)
Control class in PQM schools	-0.027	0.056	-0.044	-0.024	0.010	-0.081
	(0.068)	(0.059)	(0.083)	(0.069)	(0.054)	(0.085)
Observations	812	932	886	812	932	886

Note. Difference-in-differences estimates of the effect of the intervention on mathematics and Italian language. The top panel refers to gains in mathematics - columns (1), (2) and (3) - and Italian language - columns (4), (5) and (6). Mathematics is decomposed into mathematical reasoning (central panel) and and mathematical knowledge (bottom panel). Italian language is decomposed into reading comprehension (central panel) and grammar knowledge (bottom panel). Schools have been divided into three groups according to test scores in pre-treatment year as explained in Section 5.2. Estimates are at the class level with *sezione* fixed effects. Standard errors clustered at the school level in parentheses. *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$

Table 7: Effect of PQM on mathematics and Italian language (Intensity)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
		Bottom Tertile				Middle Tertile				Top Tertile		
	Average	25^{th}	50^{th}	75^{th}	Average	25^{th}	50^{th}	75^{th}	Average	25^{th}	50^{th}	75^{th}
Panel A: Mathematics												
Any extra class mathematics	0.302**	0.064	0.127***	0.129***	0.054	0.001	0.001	0.022	0.046	0.022	-0.010	-0.014
	(0.119)	(0.045)	(0.046)	(0.042)	(0.088)	(0.032)	(0.045)	(0.035)	(0.091)	(0.024)	(0.043)	(0.043)
Intensity mathematics	-0.001	-0.002	-0.000	0.001	0.000	0.001	0.002	-0.000	0.009*	0.004***	0.004*	0.004
	(0.006)	(0.003)	(0.003)	(0.002)	(0.004)	(0.002)	(0.002)	(0.002)	(0.005)	(0.001)	(0.002)	(0.003)
Any extra class language	-0.015	0.049	-0.024	-0.011	-0.044	-0.013	-0.013	-0.032	-0.292**	-0.044	-0.089*	-0.152***
	(0.085)	(0.038)	(0.038)	(0.027)	(0.094)	(0.032)	(0.043)	(0.038)	(0.124)	(0.028)	(0.051)	(0.052)
Intensity language	-0.006	-0.001	-0.001	-0.003	-0.007	-0.002	-0.003	-0.002	0.004	-0.001	0.005	-0.002
	(0.007)	(0.002)	(0.003)	(0.003)	(0.005)	(0.002)	(0.003)	(0.002)	(0.009)	(0.003)	(0.004)	(0.004)
Control class in PQM school	0.016	-0.005	0.023	0.000	0.034	0.039	0.021	0.004	-0.033	0.004	-0.005	-0.026
	(0.069)	(0.030)	(0.030)	(0.022)	(0.071)	(0.026)	(0.033)	(0.028)	(0.089)	(0.021)	(0.034)	(0.036)
Panel B: Italian Language												
Any extra class mathematics	0.003	0.009	0.013	0.003	0.074	0.016	0.040	0.058*	0.105	0.053**	0.024	0.013
	(0.098)	(0.041)	(0.042)	(0.029)	(0.078)	(0.028)	(0.033)	(0.030)	(0.083)	(0.026)	(0.041)	(0.034)
Intensity mathematics	-0.004	-0.002	0.001	-0.000	-0.004*	-0.001	-0.004***	-0.001	0.012***	0.005***	0.007***	0.005***
	(0.007)	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.002)	(0.002)
Any extra class language	-0.115	-0.019	-0.042	-0.036	0.042	0.003	0.025	0.004	0.036	0.007	0.027	0.004
	(0.082)	(0.033)	(0.033)	(0.025)	(0.085)	(0.030)	(0.040)	(0.031)	(0.086)	(0.027)	(0.043)	(0.039)
Intensity language	-0.013	-0.005	-0.006	-0.004	-0.000	-0.001	0.001	0.002	0.027***	0.008***	0.012**	0.008**
	(0.009)	(0.003)	(0.004)	(0.003)	(0.006)	(0.003)	(0.002)	(0.002)	(0.006)	(0.002)	(0.005)	(0.003)
Control class in PQM school	-0.054	-0.016	-0.025	-0.010	0.019	0.023	0.010	-0.030*	-0.070	-0.019	-0.051	-0.025
	(0.068)	(0.027)	(0.029)	(0.020)	(0.054)	(0.024)	(0.023)	(0.018)	(0.084)	(0.026)	(0.036)	(0.033)
Observations	806	806	806	806	920	920	920	920	880	880	880	880

Note. Difference-in-differences estimates of the effect of the intensity of the intervention on mathematics and Italian language. Estimates are at the class level with *sezione* fixed effects. Standard errors clustered at the school level in parentheses. For 6 classes enrolled in PQM mathematics and 6 classes enrolled in PQM Italian language we do not have information about the number of activities taken during the afternoon; thus those classes are not used in computations. *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$

Table 8: Robustness checks (variability of test scores explained by *sezione*)

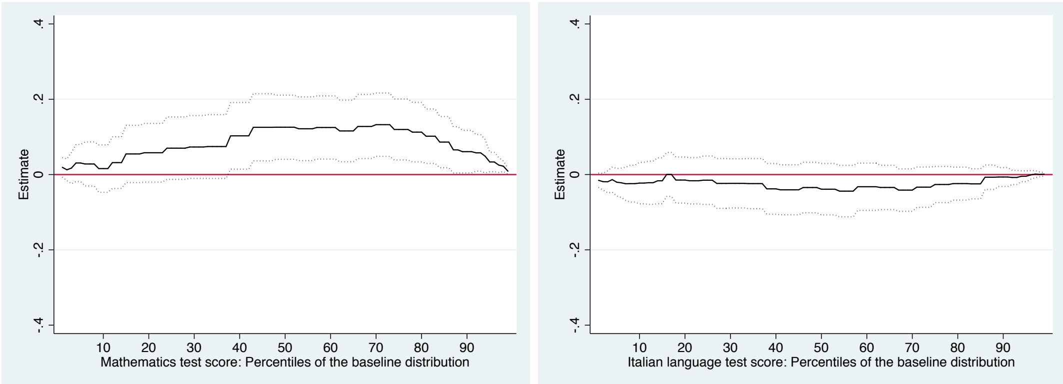
A: % of variance explained by school, *sezione* and class (pooled 2009/10 and 2010/11 data)			
	School	Class	Residual
Mathematics	0.113	0.208	0.696
Italian Language	0.104	0.143	0.758
	School	*Sezione*	Residual
Mathematics	0.105	0.123	0.783
Italian Language	0.099	0.087	0.815
B: Serial correlation of variables in the same *sezione* across years			
	Null Model	School FE	School FE + controls
Test score mathematics	0.419***	0.265***	0.230***
	(0.011)	(0.014)	(0.014)
Test score language	0.478***	0.282***	0.223***
	(0.011)	(0.014)	(0.014)
Home possession coefficient	0.327***	0.150***	0.088***
	(0.012)	(0.014)	(0.014)
% of foreign students	0.263***	0.151***	0.136***
	(0.012)	(0.014)	(0.014)

Note. Panel A. Multilevel model that makes use of school and class membership (first row), and school and *sezione* membership (second row); all sixth graders in the four PON regions are considered, excluding PQM schools. Panel B. Regression of variables at time 2 (2010/11) on variable at time 1 (2009/10) in the same *sezione* (null model); including school fixed effects (FE); and adding controls - see Section 8.2 for details. *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$

Figure 1: Effects on quantiles of test scores (within class variability)

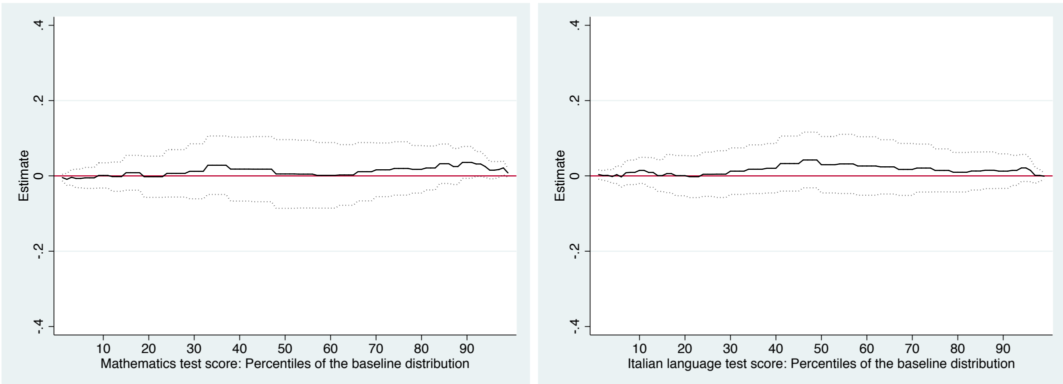
Bottom tertile

Mathematics | Italian Language



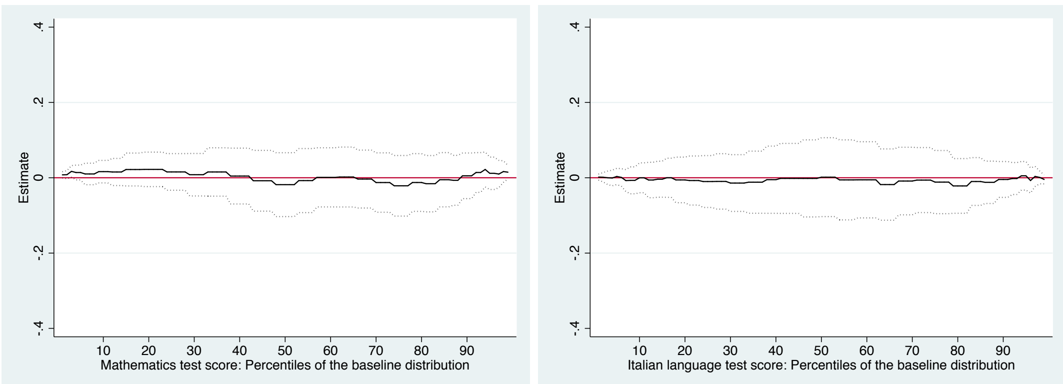
Middle tertile

Mathematics | Italian Language



Top tertile

Mathematics | Italian Language

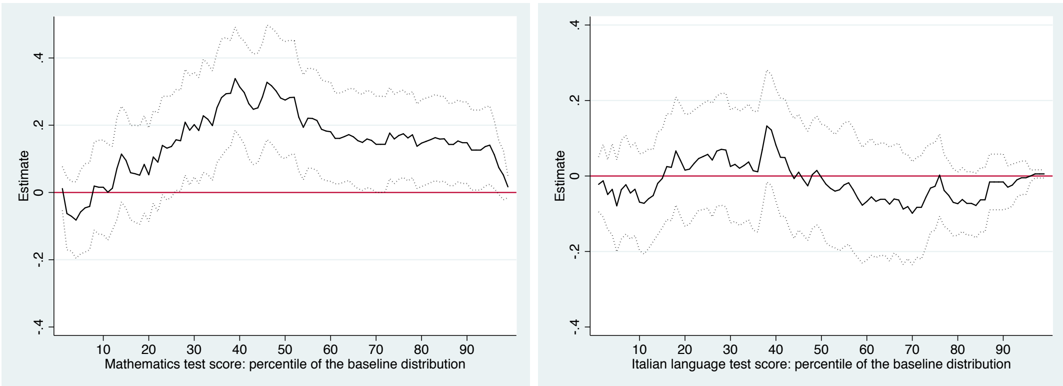


Note. Effects of PQM within classes for test scores in mathematics (first column) and Italian language (second column), together with 95% confidence intervals. Point estimates were derived by estimating equation (1), where the outcome variable is the percentage of students in the class with test score above the percentile reported on the horizontal axis. Percentiles were computed from pre-programme data. Confidence intervals were obtained using a normal approximation, standard errors being clustered at the school level. See Section 6 for details.

Figure 2: Effects on quantiles of test scores (between class variability).

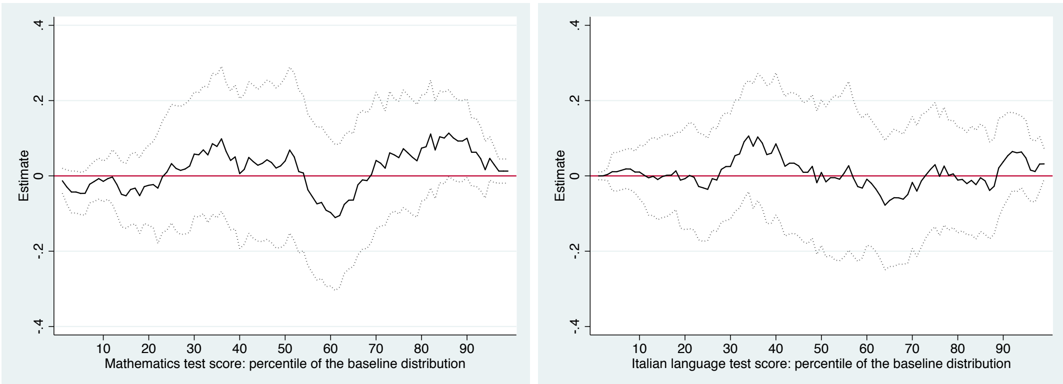
Bottom tertile

Mathematics | Italian Language



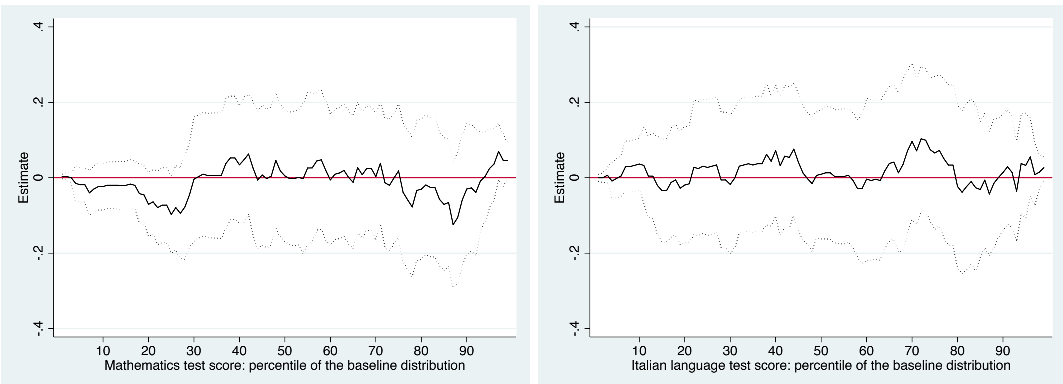
Middle tertile

Mathematics | Italian Language



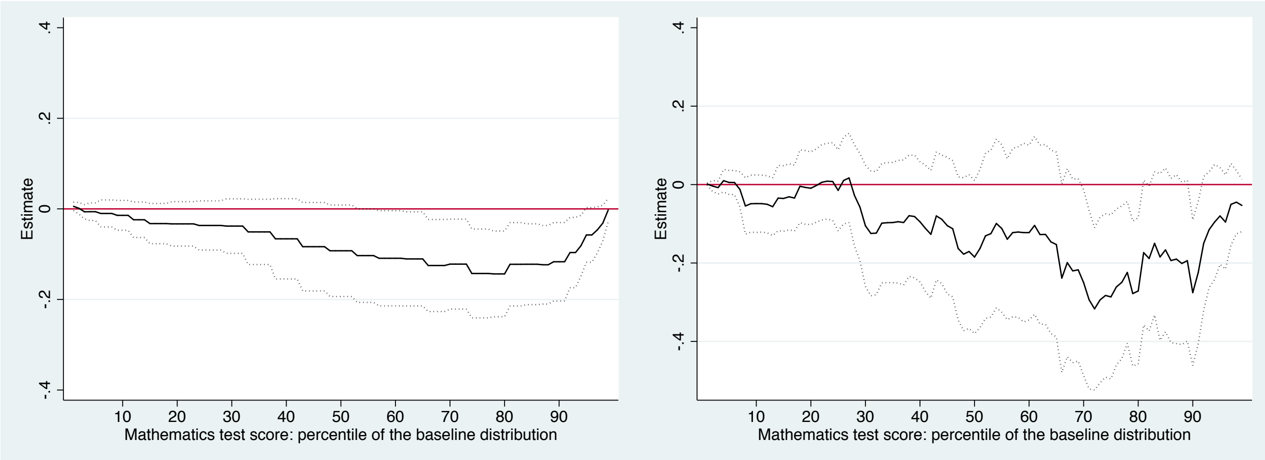
Top tertile

Mathematics | Italian Language



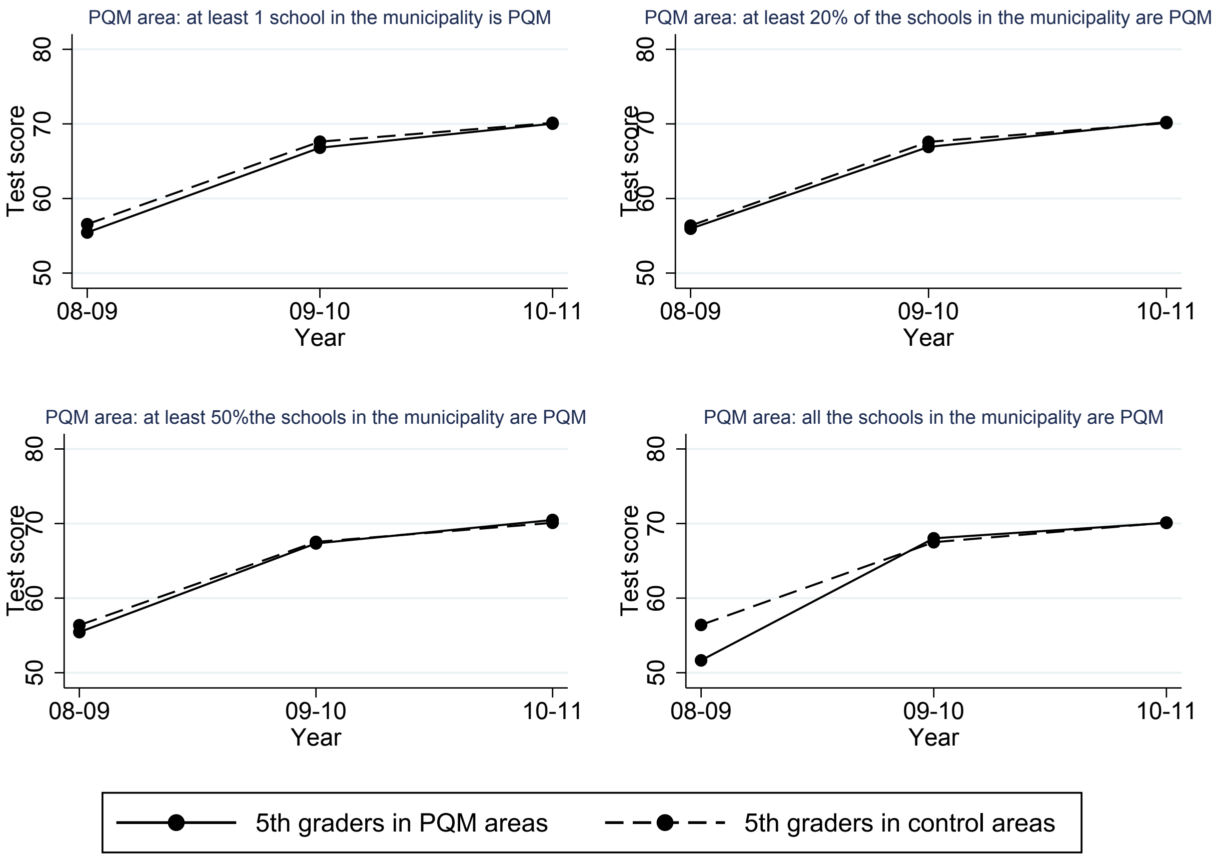
Note. Effects of PQM across classes for test scores in mathematics (first column) and Italian language (second column), together with 95% confidence intervals. Point estimates were derived by estimating equation (1), where the outcome variable is a dummy for the class average score being above the percentile reported on the horizontal axis. Percentiles were computed from pre-programme data. Confidence intervals were obtained using a normal approximation, standard errors being clustered at the school level. See Section 6 for details.

Figure 3: Effects of Italian language extra instruction time on quantiles of mathematics test scores (between and within class variability)



Note. Effects of PQM Italian language within (left hand side panel) and between (right hand side panel) classes for test scores in mathematics, together with 95% confidence intervals, for schools in the top tertile. Point estimates were derived by estimating equation (1) as explained in the text. Confidence intervals were obtained using a normal approximation, standard errors being clustered at the school level. See Section 6 for details.

Figure 4: Specification test using data for fifth graders (mathematics)



Note. The analysis uses all municipalities with at least one lower secondary school. We consider 955 municipalities for the years 2009/10 and 2010/11 917 (112 of which have at least one PQM school), and 135 municipalities for the year 2008/09 (30 with at least one PQM school).

Appendix A: Additional Results (for on-line publication)

Figure A-1: Number and location of PQM and control schools in the different municipalities of the four Objective 1 regions

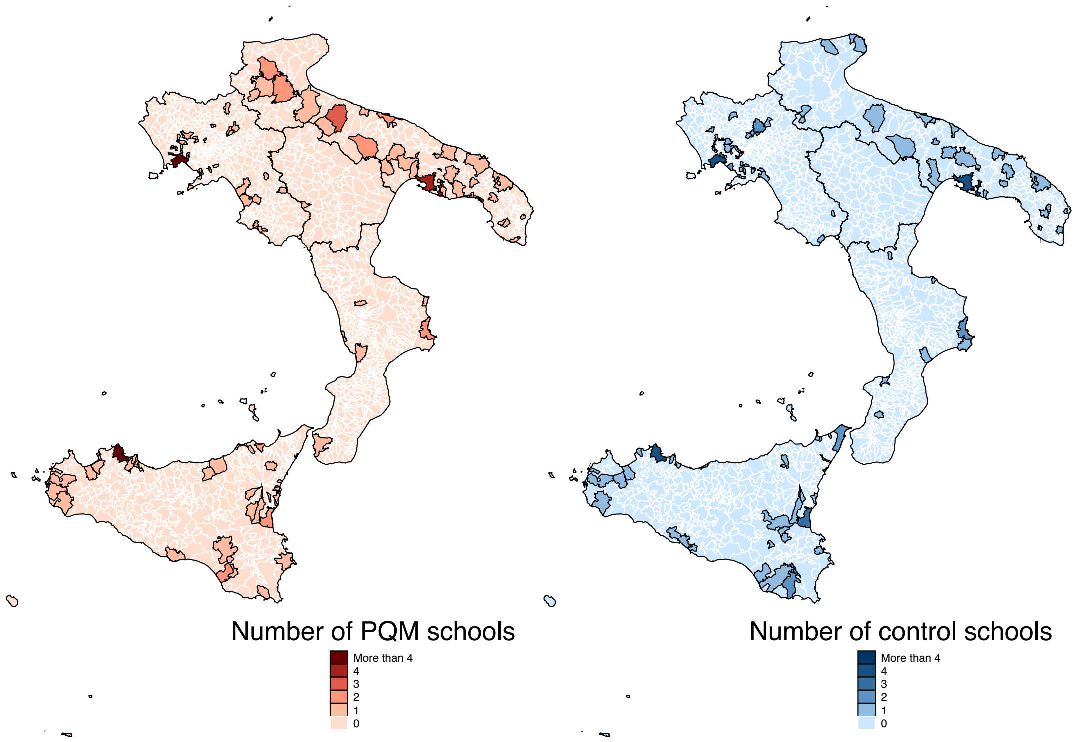


Table A-1: Descriptive of student average characteristics in treated and non treated school

	PQM schools	Control schools	Difference
Test score mathematics	-0.035	0.010	-0.044
			(0.055)
Test score language	-0.067	-0.001	-0.066
			(0.050)
% of female students	0.494	0.492	0.002
			(0.008)
% of ahead students	0.026	0.030	-0.004
			(0.004)
% of behind students	0.060	0.054	0.005
			(0.006)
% of immigrant students	0.034	0.032	0.002
			(0.004)
% of mothers employed	0.362	0.422	-0.060
			(0.021)
% of students whose father's occupation is : unemployed	0.071	0.052	0.018
			(0.009)
% of students whose father's occupation is : blue collar	0.307	0.304	0.003
			(0.019)
% of students whose father's occupation is : white collar	0.425	0.425	0.000
			(0.017)
% of students whose father's occupation is : managerial	0.197	0.218	-0.021
			(0.016)
% of students whose parents have low education	0.459	0.407	0.052
			(0.027)
% of students whose parents have medium education	0.399	0.415	-0.017
			(0.018)
% of students whose parents have high education	0.142	0.177	-0.035
			(0.018)
Average HOME scale coefficient	-0.065	-0.049	-0.016
			(0.032)
% of students living with both parents	0.900	0.892	0.008
			(0.006)
% of student attending childcare before age 3	0.278	0.270	0.008
			(0.034)
Class average weekly hour	31.754	31.418	0.336
			(0.316)
Average class size	21.962	22.431	-0.469
			(0.385)
Parents' education missing variable	0.234	0.293	-0.059
			(0.041)
Father work missing variable	0.219	0.260	-0.041
			(0.040)
Mother work missing variable	0.189	0.228	-0.039
			(0.041)
Number schools	134	114	

Note. Presented are descriptive statistics at the student level for schools in the working sample (pre-programme data only). Column (1) and (2): mean value of the variables in PQM and control schools; column (3): difference between column (1) and column (2) - standard errors in parentheses.

Table A-2: Descriptive statistics (determinants of programme intensity)

	Bottom tertile		Middle tertile		Top tertile	
	(1)	(2)	(3)	(4)	(5)	(6)
	Intensity math	Intensity language	Intensity math	Intensity language	Intensity math	Intensity language
Standardised % of correct answers in mathematics	-0.050	0.002	0.012	0.030	0.045^{**}	0.010
	(0.036)	(0.011)	(0.010)	(0.026)	(0.022)	(0.029)
Standardised % of correct answers in language	-0.055	0.005	-0.002	-0.033	-0.066^{*}	0.001
	(0.047)	(0.013)	(0.008)	(0.022)	(0.039)	(0.048)
% of female students	0.078	0.024	-0.021	0.003	-0.058	-0.002
	(0.085)	(0.020)	(0.025)	(0.044)	(0.048)	(0.050)
% of ahead students	0.095	-0.031	-0.137	-0.061	-0.340	-0.065
	(0.231)	(0.086)	(0.211)	(0.172)	(0.277)	(0.190)
% of behind students	-0.303	0.007	0.071	-0.194	-0.269	-0.111
	(0.251)	(0.051)	(0.103)	(0.125)	(0.264)	(0.114)
% of foreign students	-0.082	0.064	-0.083	0.185	0.234	-0.164
	(0.304)	(0.068)	(0.073)	(0.271)	(0.149)	(0.124)
Class weekly hour	-0.010	-0.003	0.003	0.001	-0.008	0.004
	(0.008)	(0.003)	(0.004)	(0.005)	(0.010)	(0.003)
Class size	0.002	0.002	0.000	0.001	0.005^{*}	-0.001
	(0.006)	(0.002)	(0.001)	(0.003)	(0.003)	(0.002)
Constant	0.580^{*}	0.261^{**}	0.369^{**}	0.210	0.567^{*}	0.124
	(0.348)	(0.103)	(0.151)	(0.198)	(0.327)	(0.126)
Observations	57	74	57	59	46	47

Note. Presented are estimates obtained by regressing intensity in mathematics and language at time 2 (2010/11) on inputs in the same *sezione* at time 1 (2009/10). Estimates are at the *sezione* level. Only PQM *sezioni* have been kept. Schools are divided into tertiles defined from the average test score in mathematics in year 2009/10 (see Section 5.2 for details).

Table A-3: Effect of PQM on mathematics and Italian language (participation vis-à-vis non participation) - Stratification based on Italian language test score

	(1) Bottom tertile	(2) Middle tertile	(3) Top tertile	(4) Bottom tertile	(5) Middle tertile	(6) Top tertile
	Mathematics			Italian language		
Any extra class in mathematics	0.148	0.146	-0.004	0.001	0.018	0.067
	(0.099)	(0.125)	(0.082)	(0.081)	(0.095)	(0.075)
Any extra class in language	-0.017	0.013	-0.237**	-0.044	-0.018	-0.021
	(0.080)	(0.088)	(0.105)	(0.081)	(0.079)	(0.076)
Control class in PQM schools	-0.049	0.078	-0.028	-0.122*	0.019	-0.030
	(0.067)	(0.073)	(0.087)	(0.071)	(0.053)	(0.077)
	Mathematical reasoning			Reading comprehension		
Any extra class in mathematics	0.173**	0.158	-0.027	-0.028	0.064	0.022
	(0.083)	(0.121)	(0.091)	(0.074)	(0.091)	(0.074)
Any extra class in language	0.012	0.005	-0.190*	-0.065	0.026	-0.060
	(0.080)	(0.089)	(0.103)	(0.073)	(0.073)	(0.079)
Control class in PQM schools	-0.031	0.081	-0.036	-0.136**	-0.014	-0.037
	(0.072)	(0.073)	(0.085)	(0.063)	(0.050)	(0.072)
	Mathematical knowledge			Grammar knowledge		
Any extra class in mathematics	0.109	0.127	0.022	0.076	-0.068	0.079
	(0.108)	(0.116)	(0.076)	(0.085)	(0.104)	(0.078)
Any extra class in language	-0.049	0.011	-0.243**	0.022	-0.050	0.044
	(0.081)	(0.084)	(0.100)	(0.084)	(0.081)	(0.074)
Control class in PQM schools	-0.068	0.059	-0.005	-0.080	0.027	-0.043
	(0.063)	(0.069)	(0.081)	(0.072)	(0.065)	(0.078)
Observations	772	892	966	772	892	966

Note. Difference-in-differences estimates of the effect of the intervention on mathematics and Italian language. The top panel refers to gains in mathematics - columns (1), (2) and (3) - and Italian language - columns (4), (5) and (6). Mathematics is decomposed into mathematical reasoning (central panel) and and mathematical knowledge (bottom panel). Italian language is decomposed into reading comprehension (central panel) and grammar knowledge (bottom panel). Schools have been divided into three groups according to test scores in language in pre-treatment year. Estimates are at the class level with *sezione* fixed effects. Standard errors clustered at the school level in parentheses. *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$

Table A-4: Change-in-change estimates of any extra class in mathematics on mathematics test score in schools belonging to the bottom tertile

	Control classes in PQM schools (1)	Control classes in control schools (2)	Control classes in PQM and control schools (3)
Average	0.301***	0.343***	0.326***
	(0.048)	(0.044)	(0.044)
25^{th}	0.141**	0.183***	0.166***
	(0.068)	(0.067)	(0.046)
50^{th}	0.360***	0.402***	0.385***
	(0.079)	(0.068)	(0.046)
75^{th}	0.571***	0.613 ***	0.596***
	(0.107)	(0.080)	(0.071)

Note. In the Table we report the Change-in-changes (Athey and Imbens, 2006) estimates of the effect of receiving any extra class in mathematics on mathematics test scores only in schools beloninging to the bottom tertile. We report average treatment effect, and quantile treatment effect on the 25^{th}, 50^{th} and 75^{th} quantile. In the CIC settings we could not deal with multiple treatments - any extra class in mathematics and in language- therefore the estimates refer to the effect of any extra class in mathematics without controlling for classes receiveing treatment also in language. We could not deal with multiple control groups either, so we run three regressions: (1) usinig as controls only control classes in PQM schools; (2) using as controls only control classes in control schools; (3) using as controls control clases in PQM and control schools, without distinghshing the two groups. The estimates reported control for covariates using a weighting procedure as presented in Meroni (2013). *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$

Table A-5: Adjusted p-values

Tertile	Regression	Coefficient	Observed (1)	Adjusted p-values (2)	(3)	(4)
Bottom	Mathematics	Extra class mathematics	0.010	0.045	0.016	0.030
		Extra class language	0.961	0.957	0.957	0.961
	Reasoning	Extra class mathematics	0.001	0.004	0.001	0.006
		Extra class language	0.884	0.957	0.880	0.961
	Knowledge	Extra class mathematics	0.049	0.151	0.060	0.099
		Extra class language	0.956	0.957	0.956	0.961
	Language	Extra class mathematics	0.831	0.927	0.819	0.996
		Extra class language	0.218	0.542	0.219	0.750
	Reading	Extra class mathematics	0.996	0.994	0.994	0.996
		Extra class language	0.250	0.583	0.268	0.750
	Grammar	Extra class mathematics	0.376	0.568	0.377	0.564
		Extra class language	0.551	0.868	0.534	0.961
Middle	Mathematics	Extra class mathematics	0.559	0.779	0.540	0.778
		Extra class language	0.485	0.809	0.511	0.780
	Reasoning	Extra class mathematics	0.300	0.625	0.310	0.778
		Extra class language	0.650	0.809	0.664	0.780
	Knowledge	Extra class mathematics	0.778	0.779	0.779	0.778
		Extra class language	0.368	0.736	0.391	0.780
	Language	Extra class mathematics	0.394	0.715	0.394	0.778
		Extra class language	0.541	0.809	0.535	0.780
	Reading	Extra class mathematics	0.119	0.371	0.124	0.713
		Extra class language	0.366	0.736	0.359	0.780
	Grammar	Extra class mathematics	0.682	0.779	0.678	0.778
		Extra class language	0.803	0.809	0.809	0.803
Top	Mathematics	Extra class mathematics	0.731	0.945	0.743	0.917
		Extra class language	0.020	0.074	0.031	0.059
	Reasoning	Extra class mathematics	0.947	0.953	0.953	0.947
		Extra class language	0.073	0.203	0.090	0.146
	Knowledge	Extra class mathematics	0.464	0.793	0.494	0.917
		Extra class language	0.010	0.051	0.014	0.059
	Language	Extra class mathematics	0.381	0.759	0.431	0.917
		Extra class language	0.844	0.905	0.854	0.909
	Reading	Extra class mathematics	0.764	0.945	0.795	0.917
		Extra class language	0.665	0.879	0.684	0.909
	Grammar	Extra class mathematics	0.323	0.729	0.374	0.917
		Extra class language	0.909	0.905	0.905	0.909

Note. In the Table we report: (1) the observed p-values; (2) adjusted p-values as in Westfall and Young (1993); (3) adjuested p-values as in Efron and Tibshirani (1993); (4) adjusted p-values as in Anderson (2008), as explained in Section 7.